Numerical Methods and Uniquness for the Canham-Helfrich Model of Biomembranes

A Thesis

Submitted to the Faculty

of

Drexel University

by

Jingmin Chen

in partial fulfillment of the

requirements for the degree

of

Doctor of Philosophy in Mathematics

May 2015

© Copyright 2015
Jingmin Chen. All Rights Reserved.

Acknowledgements

I would like to express my special appreciation and thanks to my adviser Professor Thomas Yu for his support, patience and encouragement during the past five years. He has provided me with not only extremely helpful guidance in completing this thesis, but also career advice to my future adventure. I also would like to thank our collaborator Professor Rob Kunser. He inspired us to solve these interesting geometric optimization problems and provided insightful discussions about the research. I also wish to thank Professor Pavel Grinfeld, Gideon Simpson and Ronald Perline for serving as my committee members. I thank Professor Justin Smith for his help in parallel computing. I am also thankful to Andrew Zigerelli for helping me develop both the matlab version and the GPU version of Subdivision Lab. I also thank Drexel University for the Dissertation Fellowship and Department of Mathematics for the financial support to my PhD study.

Table of Contents

Abstract

The classical Canham-Helfrich models of biomembranes consist of a family of geometric constrained variational problems. Their physical importance and mathematical challenge attract the attention of both biophysicists and geometric analysts. In this PhD thesis, we develop a numerical method for these models. Our method uses a high-order approximation of surfaces with arbitrary topology based on subdivision methods. We also develop multiscale and parallel versions of our method which substantially speed up computations. An implementation based on Matlab and CUDA is provided along with this thesis. We use our solver to explore a phenomenon known as *conformal diffusion* in the biophysical literature, which is also connected to the open uniqueness question for the Canham and Helfrich variation problems. We establish the uniqueness of solution of the Canham problem in a special case related to the Willmore conjecture (now the Marques-Neves theorem).

1. Introduction

Lipid bilayers are the most elementary and indispensable structural components of biological membranes, which form the boundary of most cells. Since lipid bilayers also exhibit a great variety of different shapes (see the collection of figures in [33]), biophysicists have strived to find good mathematical models for such biomembranes. It is known since the seminal work of Canham [9], Helfrich [17] and Evans [15] in the 70's that elastic bending energy, subject to volume and area and possibly other constraints, plays the key role in driving the geometric configurations of such membranes. In contrast to isotropic surface tension in simple liquids where energy minimization leads to round spherical droplets only, the curvature-based bending energy explains the variety of shapes observed, even in closed vesicles with a spherical topology.

The so-called bilayer couple (BC) model (referred to as the Helfrich-BC model, or simply the **Helfrich model**) suggests that a biomembrane surface S configures itself so as to minimize $\int_S H^2 dA$ subject to the area, volume and area difference (related to the bilayer characteristics) constraints. It is observed experimentally that no topological change occurs in any accessible time-scale, so, by Gauss-Bonnet, $\int_S K dA$, being a topological invariant, remains fixed, and we can restate the variational problem as: Given a genus g and admissible values of (A_0, V_0, M_0) (subject to isoperimetric

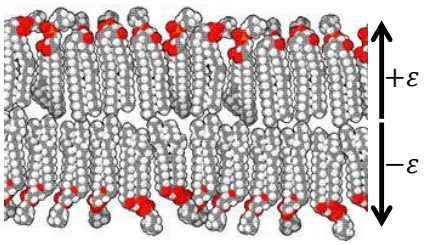


Figure 1.1: The 'offset surfaces' of the pivotal surface of a lipid bilayer

constraints), solve

$$\min_S \overbrace{\int_S \underbrace{H^2 - K}_{(\kappa_1-\kappa_2)^2/4}\, dA}^{=:\text{Willmore}(S)} \text{ s.t. } \begin{cases} \text{(i)} & \text{area}(S) = \int_S 1\, dA = A_0, \\ \text{(ii)} & \text{volume}(S) = \frac{1}{3}\int_S [x\hat{\mathbf{i}} + y\hat{\mathbf{j}} + z\hat{\mathbf{k}}] \cdot \hat{\mathbf{n}}\, dA = V_0, \\ \text{(iii)} & \text{total-mean-curvature}(S) := \int_S H\, dA = M_0, \end{cases} \tag{1.1}$$

over all genus g surfaces S. Here $H = (\kappa_1 + \kappa_2)/2$ and $K = \kappa_1\kappa_2$ are the mean and Gauss curvatures. In (ii), we use the divergence theorem to express the volume enclosed by S as a surface integral. In (iii), we use the relation $\int_S H dA = \lim_{\varepsilon\to 0} \frac{1}{4\varepsilon}(\text{area}(S_{+\varepsilon}) - \text{area}(S_{-\varepsilon}))$, where $S_{+\varepsilon}$ and $S_{-\varepsilon}$ are the 'ε-offset surfaces' (see Figure 1.1), and exploit the fact that the thickness of the lipid bilayer, 2ε, is negligible compared to the size of the vesicle. The constraint values A_0, V_0 and M_0 are determined by physical conditions, and are subject to the isoperimetric constraints.

The objective function Willmore(S) in the variational problem (1.1) is called the **Willmore energy** of the surface S. When the area-difference constraint (iii) in (1.1) is omitted, the model is referred to as the **Canham model**. The existence of a minimizer for the Canham model was recently established in [32] for the genus 0 case and in [21] for arbitrary genus.

It is well-known from [3, 11] that the quantity $(H^2 - K)dA$ is invariant under Möbius transformations, i.e. any transformation from the group of translations (3 dimensions), rotations (3 dimensions), uniform scalings and ***sphere inversions*** (1+3=4 dimensions). If we denote this (Lie) group by Möb(3); we have $\dim(\text{Möb}(3)) = 3 + 3 + 4 = 10$. This dimensionality count will play an interesting role in our study.

The mathematical and numerical analysis of Helfrich's variational problem are not well-studied. A number of biophysicists and applied mathematicians developed numerical methods for solving the Canham or Helfrich model; see [18, 8, 26, 4, 5, 7]. It

is understandable that none of these numerical methods offer a convergence guarantee, as even the existence and uniqueness question of the Helfrich's problem is unsettled.

There have been a few recent theoretical breakthroughs, most notably the solution of the Willmore conjecture [24]. It will help our numerical work by offering some background of these recent results and the related open problems.

1.1 Background

The **Willmore conjecture**, a pure geometry problem, is quite related to the Helfrich and Canham models: it states that among all closed genus 1 (i.e. toroidal) surfaces, the torus with (major radius)/(minor radius) $= \sqrt{2}$, which we refer to as the *symmetric Clifford torus* or just *Clifford torus*,[1] has the lowest Willmore energy. In other words, the Clifford torus solves the unconstrained version of Helfrich's problem in the genus 1 case. Posed by geometer Tom Willmore in 1965, and predating the biophysical works of Canham [9], Helfrich [17] and Evans [15], the problem attracted the attention of many geometric analysts. The Willmore conjecture was solved by Marques and Neves in 2012 [24].

Notice from the definition of Willmore(S) in (1.1) that if S is the round sphere, then $\kappa_1 = \kappa_2$ everywhere and therefore Willmore$(S) = 0$. Conversely, any S with Willmore$(S) = 0$ is totally umbilic, i.e. round. So the genus 0 version of Willmore's problem is almost trivial. To find a genus 1 surface with low Willmore energy, one may think that a good candidate would be "a round sphere with a very small handle", and by making the handle smaller and smaller the genus 1 surface can have a Willmore

[1] In geometric analysis, the term 'Clifford torus' refers to

$$S^1(1/\sqrt{2}) \times S^1(1/\sqrt{2}) = \left\{(\cos(\theta), \sin(\theta), \cos(\phi), \sin(\phi)/\sqrt{2} : \theta, \phi \in [0, 2\pi]\right\} \subset S^3 \subset \mathbb{R}^4.$$

By applying a suitable stereographic projection we can map this surface in 4-dimension to what we call the 'symmetric Clifford torus' in three-dimension. However, with different choices of stereographic projections (or by fixing the stereographic projection but letting $O(4)$ act on the Clifford torus sitting in S^3), we get instead Dupin cyclides; we refer to the latter as 'Clifford cyclides'.

energy arbitrarily close to that of a round sphere. But the Möbius invariance of Willmore energy immediately falsifies this intuition: under a suitable sphere inversion, a very small handle becomes a very big handle in the Euclidean metric! It is for this reason that even the *existence* of a genus 1 minimizer is a nontrivial problem[2]. For the Canham model, the existence problem was recently resolved in [32] for the genus 0 case and in [21] for arbitrary genus, at least for a range of constraint isoperimetric values.

Equally intriguing is the *uniqueness* question, for both mathematical and biophysical reasons:

- Since all four functionals (Willmore energy, area, volume, bilayer area difference) in the Helfrich model are invariant under rigid transformations, by uniqueness we of course mean uniqueness up to rigid transformation. Recall that the Willmore energy is invariant under the 10-dimensional group Möb(3), which contains the 6-dimensional subgroup of rigid motion; this leaves us 4 (non-rigid) degrees of freedom against the 2 constraints in the Canham model, or 3 constraints in the Helfrich model. Based on this count, one expects a two-parameter family of solutions in the Canham model and a one-parameter family for the Helfrich model.

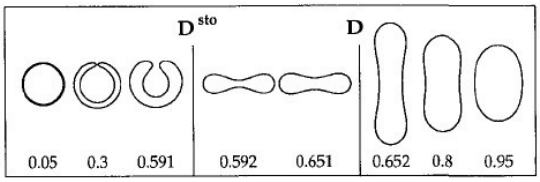


Figure 1.2: Conjectured unique surface of revolution solutions of the Canham model with different reduced volume $v_0 := \frac{3V_0}{4\pi}(\frac{A_0}{4\pi})^{-\frac{3}{2}}$.

[2]First solved by L. Simon, in [34].

However, there is a glitch in this dimensionality count: what if the postulated one or two-parameter family of Möbius transformations all happen to be (self-)diffeomorphisms of a unique solution surface? For instance, if we set the constraint values (A_0, V_0, M_0) in (1.1) to be $(4\pi, 4\pi/3, 4\pi)$, i.e. the corresponding values of a round unit sphere, then we know that any unit sphere solves (1.1). In this very case, no Möbius transformation can generate a new solution, because Möbius transformations map spheres to spheres. Indeed, we observe from computer experiments that, in the genus 0 case, uniqueness appears to hold for all admissible values of (A_0, V_0, M_0); moreover the unique minimizer always appears to be a *surface of revolution*; see Figure 1.2. We believe that these two observations are linked.

- On the other hand, the dimensionality count works when the genus is 2 or higher. It is because *Möbius transformations are conformal*. By Hurwitz's automorphisms theorem, when $g \geqslant 2$, there can at most be $84(g-1)$, in particular finite, conformal automorphisms on a genus g surface. In other words, there cannot be an infinite 1- or 2-parameter family of Mobius automorphisms on any genus 2 surface. So the glitch mentioned in the previous paragraph can only come into play when $g = 0$ or 1. In particular, we expect non-uniquness for vesciles with genus 2 or above. Such non-uniquness in genus 2 lipid bilayer vesicles are observed by biophysicists under their microscopes [20, 22]. Biophysicists believe that this non-uniqueness explains the peculiar shape variation observed experimentally for vesicles with genus $g \geqslant 2$, and call this phenomenon "conformal diffusion". These discoveries give confidence to the accuracy of Helfrich's model.

1.2 This paper

Motivated by the intertwining mathematical and biophysical interests, we are interested in developing efficient numerical methods for the Helfrich problem. We propose a numerical method which is *high order*, *multiscale* and *parallelizable*. In contrast to previous methods [18, 8, 26, 4, 5, 7], the approximation space of surfaces used in our method is based on subdivision surfaces, which are smooth enough to have square integrable curvatures [31]. As such, our method is based on a *direct* high order numerical computation of the *true* Willmore energy. This feature implies that the method is consistent with the continuous problem; because of this we speculate that it is possible to come up with a rigorous convergence proof of the method. In this regard, the approximation results of subdivision surfaces by Arden [1], together with the recent existence results [32, 21], will be relevant.

Modelling of smooth surfaces of arbitrary topology is a well-studied problem in the CAGD community. From interactions with many researchers in applied mathematics, we feel that the difficulty of the problem and the advance made in the CAGD community are not well-understood by the general applied mathematics community. In the past decade, however, we saw an increasing awareness of combining CAGD techniques with techniques from finite element methods when solving engineering problems, most notably in the development of isogeometric analysis [12, 19]. A secondary motivation of this paper is to bring the method of subdivision surfaces to bear on scientific computing problems.

2. Numerical Solution based on Subdivision Surfaces

In this section, we propose a numerical solution of the Canham and Helfrich models based on subdivision surfaces. For simplicity, we focus ourselves on the use of Loop subdivision surfaces [23].

A Loop subdivision surface is specified by a control mesh $M = (\mathcal{V}, \mathcal{F})$ where $\mathcal{V} \in \mathbb{R}^{\#V \times 3}$ records the 3-D coordinates of the vertices of the control mesh, $\#V$ denotes the total number of vertices, and $\mathcal{F} \in \mathcal{I}^{\#F \times 3}$ is a list of triplets of indices from $\mathcal{I} := \{1, \ldots, \#V\}$ which records the bounding vertices of each of the $\#F$ triangle faces in the mesh M. We assume that the mesh realizes a closed simplicial surface. See Figure 2.1(a) for a simple genus 0 control mesh, and Figure 2.1(b) for the corresponding Loop subdivision surface. For each face f in M, there is a corresponding surface patch; see Figure 2.1(b).

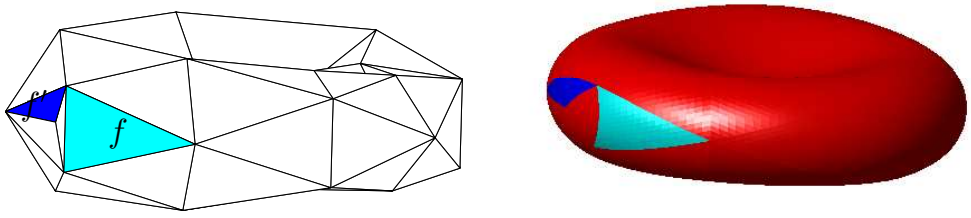


Figure 2.1: (a) A (genus 0) control mesh M; f is a regular face, f' is an irregular face (b) The Loop subdivision surface corresponding to M; in light blue: the regular patch associated to f, in deep blue: the irregular patch associated to f'

In the mono-scale version of our proposed numerical method, we assume that $\mathcal{F}$ is fixed and $\mathcal{V}$ varies.[1] For most $\mathcal{V}$, a smooth (immersed) surface, denoted by $S[\mathcal{V}]$,

[1] For the multiscale version, see Section 4.1.

is defined by the Loop scheme. Our proposed numerical method is to approximate Helfrich's variational problem (1.1) by the finite-dimensional analog:

$$\min_{\mathcal{V}\in\mathbb{R}^{\#V\times 3}} \overbrace{\int_{S[\mathcal{V}]} H^2\, dA}^{=:W(\mathcal{V})} \text{ s.t. } \begin{cases} \text{(i)} & A(\mathcal{V}) := \int_{S[\mathcal{V}]} 1\, dA = a_0 \\ \text{(ii)} & V(\mathcal{V}) := \frac{1}{3}\int_{S[\mathcal{V}]} [x\hat{\mathbf{i}} + y\hat{\mathbf{j}} + z\hat{\mathbf{k}}]\cdot\hat{\mathbf{n}}\, dA = v_0 \\ \text{(iii)} & M(\mathcal{V}) := \int_{S[\mathcal{V}]} H\, dA = m_0 \end{cases} . \tag{2.1}$$

Note that we leave out the '$-K$' term in the definition of $W(\mathcal{V})$ (compare with (1.1)) because we do not assume the topology to change.[2] In order to use a standard nonlinear optimization solvers to solve this problem, we need to develop an efficient way to compute the functionals $W(\mathcal{V})$, $A(\mathcal{V})$, $V(\mathcal{V})$, $M(\mathcal{V})$ and their corresponding gradient vectors. We do so in Section 3.

2.1 Choice of nonlinear optimization solver

We use the function `fmincon()` in the Matlab optimization toolbox Version 7.0 (R2014a). This extensive solver implements four major algorithms for constrained optimization problem: 'trust region reflective', 'active set', 'SQP', 'interior point'. A standard reference for these methods is [27]. Among these, 'interior point' is the choice for large, sparse problems, hence is the most suitable for our problem. (Note: each control vertex of a subdivision surface only affects the surface locally, hence the sparsity.)

The label 'interior point', however, is misleading here. Since our problems have only equality constraints, concepts related to interior point methods and implemented in the `fmincon` interior point solver, such as log-barrier functions or slack variables, are irrelevant to us. While the documentation is not clear about certain details, it is apparent that the solver uses a Lagrange multiplier method in conjunction with a

[2] In the rest of the paper, the term Willmore energy refers to $\int H^2 dA$ instead of $\int H^2 - K dA$.

BFGS quasi-Newton solver.[3] The Lagrange multiplier method seeks a local minimizer of the Helfrich problem (2.1) by solving the nonlinear system of equations of size $\#V + 3$ equations and $\#V + 3$ unknowns

$$\begin{aligned} \nabla W(\mathcal{V}) + \lambda_a \nabla A(\mathcal{V}) + \lambda_v \nabla V(\mathcal{V}) + \lambda_m \nabla M(\mathcal{V}) = 0 \\ A(\mathcal{V}) = a_0, \quad V(\mathcal{V}) = v_0, \quad M(\mathcal{V}) = m_0, \end{aligned} \tag{2.2}$$

where the control vertices $\mathcal{V}$ and the Lagrange multipliers λ_a, λ_v, λ_m are treated as unknowns.[4] If one were to solve this nonlinear system using a Newton method, one solves the following linear system in each Newton update:

$$\begin{bmatrix} H & \nabla A & \nabla V & \nabla M \\ \nabla A^T & 0 & 0 & 0 \\ \nabla V^T & 0 & 0 & 0 \\ \nabla M^T & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \delta\mathcal{V}^{(k)} \\ \delta\lambda_a^{(k)} \\ \delta\lambda_v^{(k)} \\ \delta\lambda_m^{(k)} \end{bmatrix} = - \begin{bmatrix} \nabla W + \lambda_a^{(k)} \nabla A + \lambda_v^{(k)} \nabla V + \lambda_m^{(k)} \nabla M \\ A - a_0 \\ V - v_0 \\ M - m_0 \end{bmatrix} \tag{2.3}$$

where all the functionals and their gradients, as well as the Hessian matrix $H := \nabla^2 W + \lambda_a^{(k)} \nabla^2 A + \lambda_v^{(k)} \nabla^2 V + \lambda_m^{(k)} \nabla^2 M$, are evaluated at the k-th Newton iterate $\mathcal{V}^{(k)}$ of vertex positions, while $\lambda_a^{(k)}, \lambda_v^{(k)}, \lambda_m^{(k)}$ are the k-th Newton iterates for the Lagrange multiplers. Since we cannot afford to compute the Hessian matrices of W, A, V, M, the

[3] The documentation mentions that the `fmincon` (interior point) solver also uses the penalty method, but the details on how it is used together with the Lagrange multiplier method are missing. There is a well-known method, called the augmented Lagrangian method, which subtly combines the idea of the Lagrange multiplier method and the penalty method; see [27, Chapter 17]. The documentation of the current version of `fmincon`, unfortunately, has no mention of the augmented Lagrangian method.

[4] These are also the so-called KKT equations of our problem, it is, of course, an overkill to use this terminology when there are no inequality constraints.

solver uses instead a quasi-Newton method based on a Broyden-Fletcher-Goldfarb-Shanno (BFGS) approximation of the Hessians.

In Section 3, we develop efficient algorithms for computing the functionals W, A, V, M and their gradients based on the Loop subdivision scheme. In the rest of this section, we review the structure of subdivision surfaces, especially the parametric description of each surface patch (recall Figure 2.1(b)) of a Loop subdivision surface.

2.2 Loop subdivision surfaces

We give an introduction to a specific subdivision surface method by Loop [23] used primarily in our optimization method. Our presentation will be brief, but contains all the necessary implementation details when read in conjunction with the paper [36] by Stam. Some references devoted to the subject of subdivision surfaces include [28, 37, 38].

Following the subdivision surface literature, a vertex is called *ordinary* if it has valence 6, otherwise it is called an *extraordinary* vertex. We assume that extraordinary vertices in M are isolated, i.e. no two extraordinary vertices can be neighbor of each other. If M lacks this property, one can simply apply a mid-point subdivision to M to resurrect that.

For each triangle face f in M, we call f a *regular* face if all its three bounding vertices are ordinary, otherwise, under our assumption, exactly one of the three vertices is extraordinary and we call f an *irregular* face. The corresponding surface patches will be simply referred to as *regular* and *irregular patches*. See again Figure 2.1.

Although the parametric description is more important for us, it is helpful (especially in Section 2.4) to recall the popular algorithmic description of a Loop surface as the limit of an iteration of subdivision steps, see Figure 2.2; the subdivision step can be succinctly described by the diagrams in Figure 2.3.

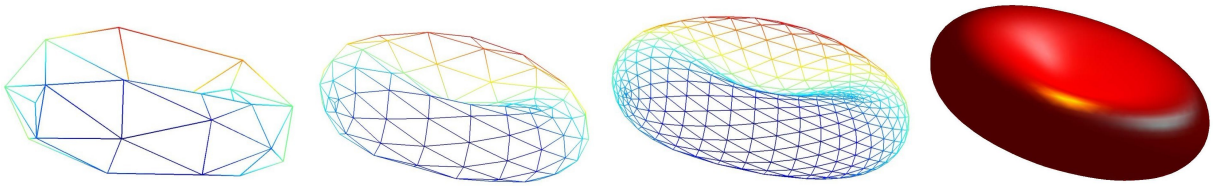


Figure 2.2: Iteration of subdivision steps

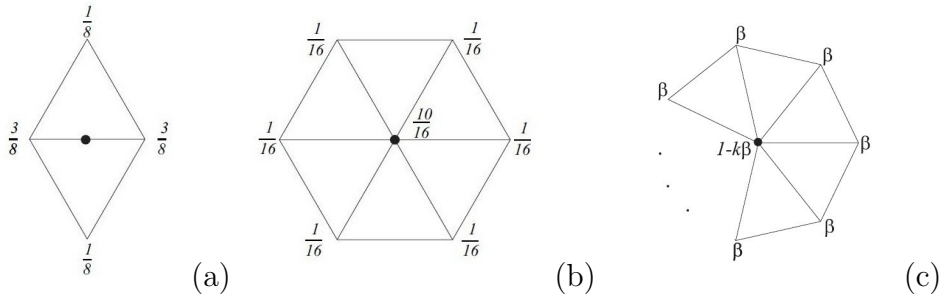


Figure 2.3: (a)The edge rule; (b) The vertex rule for an ordinary vertex; (c) The vertex rule for an extraordinary vertex, where k is the valence and $\beta = \frac{1}{k}(\frac{5}{8} - (\frac{3}{8} + \frac{1}{4}\cos(\frac{2\pi}{k}))^2)$. (In the text, we use '$N$' to denote the valence of an extraordinary vertex, so as to be consistent with Stam's paper.)

2.3 Parametrization of a regular patch

The surface patch associated with a regular face f can be parametrized by a linear combination of 12 polynomials with coefficients $\{\mathbf{c}_{f,i}\}_{i=1}^{12}$ being the coordinates of the vertices in f and their immediate neighbors ordered as in Figure 2.4(b).[5](see Figure 2.4)

$$\mathbb{R}^3 \leftarrow \Omega : \mathbf{s}_f(v,w) = \sum_{i=1}^{12} \mathbf{c}_{f,i}\mathbf{b}_i(v,w) \tag{2.4}$$

[5]When f is close to an extraordinary vertex, it is possible that some of these 12 control vertices coalesce. In such a degenerate situation, one simply repeats the coalescing vertices when using the formula (2.4).

where $\Omega := \{(v, w) : v \in [0, 1] \text{ and } w \in [0, 1 - v]\}$, and $\mathbf{b}_1(v, w), \ldots, \mathbf{b}_{12}(v, w)$ are the twelve degree 4 polynomials as shown in [36, Page 10-11]. (We do not copy these polynomials from Stam's paper, but mention that they come from the so-called M_{222} box-spline.) For notational convenience, we organize each $\mathbf{c}_{f,i}$ as a row vector of

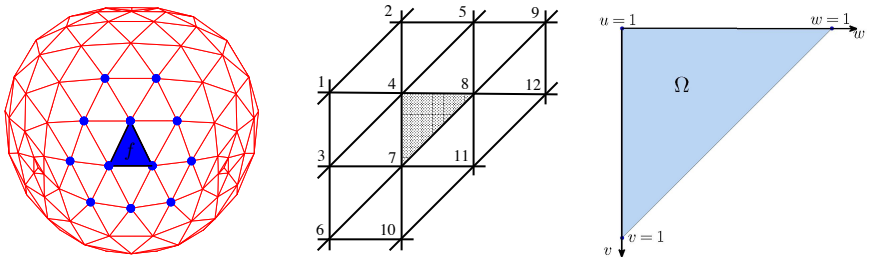


Figure 2.4: (a) A regular face and its neighboring vertices (b) ordering of the twelve control vertices (c) the parameter domain Ω. Note: for any point $(v, w) \in \Omega$, its barycentric coordinates w.r.t. to the bounding vertices of Ω, listed in the order $(0, 0)$, $(1, 0)$, $(0, 1)$, are simply $(1 - v - w, v, w)$.

length 3, write $\mathbf{c}_f := [\mathbf{c}_{f,1}; \ldots; \mathbf{c}_{f,12}] \in \mathbb{R}^{12\times 3}$, and define $\mathbf{b} := \mathbf{b}^6 := [\mathbf{b}_1; \ldots; \mathbf{b}_{12}]$ as a column vector of functions of length 12. Then (2.4) simplifies to

$$s_f = \mathbf{c}_f^T \mathbf{b}^6. \tag{2.5}$$

2.4 Parametrization of an irregular patch

The surface patch associated with an irregular face f admits a parametrization $\mathbf{s}_f : \Omega \to \mathbb{R}^3$ which is controlled by $N + 6$ control vertices around the face f, where N is the valence of the extraordinary vertex of f. Following Stam's convention, these $N + 6$ vertices are ordered as in Figure 2.5(a). Like the regular case, $\mathbf{s}_f$ is *linearly*

related to the control vertices, so

$$\mathbf{s}_f(v,w) = \sum_{i=1}^{N+6} \mathbf{c}_{f,i}\mathbf{b}_i^N(v,w) \tag{2.6}$$

for some basis functions $\mathbf{b}_i^N$, $i = 1,\ldots,N+6$, implicitly defined by the subdivision process. Unlike the regular ($N = 6$) case, none of these basis functions is a single polynomial anymore. Instead, it is an infinite piecewise polynomial, with pieces being the (recursively defined) sub-triangles Ω_k^j, $j = 1,2,\ldots$, $k = 1,2,3$, as shown in Figure 2.5(b). Note that in this figure the origin $(0,0)$ corresponds to the extraordinary vertex of f.

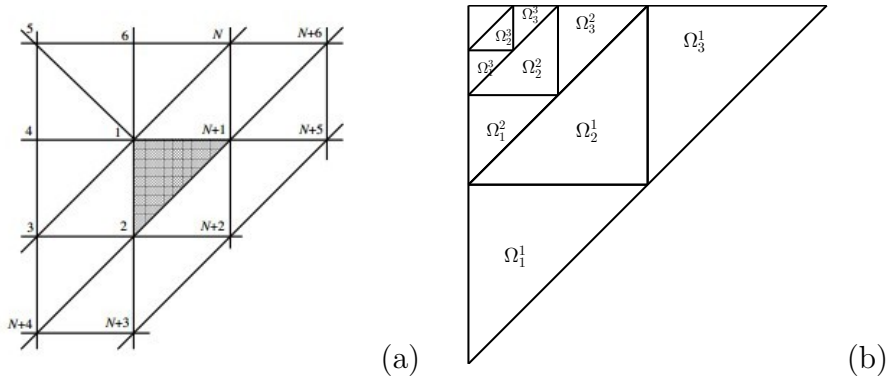


Figure 2.5: (a) Ordering of the $N+6$ control vertices around an irregular face (b) Partition of the parameter domain Ω

The parametrization (2.6) is tricky to compute; and this is where Stam's idea [36, 35] comes in. As in the regular case, write $\mathbf{c}_f := [\mathbf{c}_{f,1};\ldots;\mathbf{c}_{f,N+6}] \in \mathbb{R}^{(N+6)\times 3}$, and $\mathbf{b}^N := [\mathbf{b}_1^N;\ldots;\mathbf{b}_{N+6}^N]$. In a nutshell, Stam's method transforms the control data

$\{\mathbf{c}_{f,i}\}$ into 'eigen- control data' $\widehat{\mathbf{c}}_f = V^{-1}\mathbf{c}_f$, so

$$\mathbf{s}_f = \mathbf{c}_f^T \mathbf{b}^N = \widehat{\mathbf{c}}_f^T \underbrace{V^T \mathbf{b}^N}_{=:\Phi} = \widehat{\mathbf{c}}_f^T \Phi. \tag{2.7}$$

Here $V \in \mathbb{R}^{(N+6)\times(N+6)}$ is the matrix of (generalized) eigenvectors of the matrix A (same notation as in Stam's paper) that maps the $N+6$ control vertices around f to $N+6$ control points in the next subdivision level as shown in Figure 2.6(b), so $AV = V\Lambda$ where Λ is in a Jordan canonical form. For the Loop scheme, Λ is diagonal when the valence N is greater than 3, but has a Jordan block of size 2 when $N = 3$. Since the subdivision process is linear and *stationary*, i.e. the same linear subdivision rules are used across different scales, recall Figure 2.2, we have

$$\mathbf{c}_f^T \mathbf{b}^N(v,w) = (A\mathbf{c}_f)^T \mathbf{b}^N(2v,2w), \quad (v,w) \in \frac{1}{2}\Omega.$$

Putting these together, we have

$$\Phi(v,w) = \Lambda^T \Phi(2v,2w), \quad (v,w) \in \frac{1}{2}\Omega. \tag{2.8}$$

The key point is that these *eigenbasis functions* $\Phi = [\phi_1; \ldots; \phi_{N+6}]$ are easier to evaluate compared to the original basis functions $\mathbf{b}^N$: When $N > 3$, $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_{N+6})$ is diagonal, and we have $\phi_i(v,w) = \lambda_i \phi(2v,2w)$, apply this recursively we have

$$\phi_i(v,w) = \lambda_i^{n-1} \phi(2^{n-1}v, 2^{n-1}w), \quad \text{when } (v,w) \in \Omega_k^n. \tag{2.9}$$

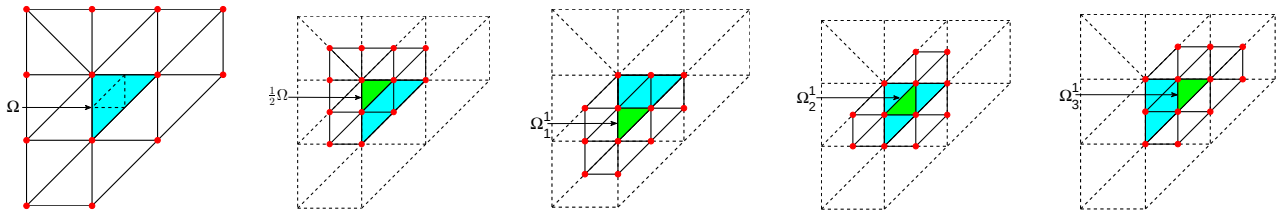
(Recall Figure 2.5(b).) As a result, each ϕ_i is specified by three – not infinitely many – polynomials. Also, in virtue of (2.9), ϕ_i and its derivatives can be easily evaluated at arbitrary parameter values after the three polynomials are specified.

These three polynomials can be evaluated based on the same polynomial basis $\mathbf{b}$

from the regular case (2.4). We write V_i as the (generalized) eigen-vector associated to the eigen-basis function ϕ_i. For $k = 1, 2, 3$, there are suitable linear maps M_k (expressed as $P_k\overline{A}$ in Stam's paper) so that M_kV_i contains the data at the 12 control vertices that determine the polynomial $\phi_i|_{\Omega_k^1}$; see Figure 2.6(c)-(e). With an appropriate affine reparametrization of Ω_k^1 by Ω, denoted as $t_{1,k} : \Omega_k^1 \to \Omega$ by Stam, this polynomial can be expressed as

$$\phi_i|_{\Omega_k^1}(v, w) = (M_kV_i)^T\mathbf{b}(t_{1,k}(v, w)). \tag{2.10}$$

See [36] for details, e.g. on how to exploit the circulant structure in the matrix A in order to facilitate the computation of the related matrices V, V^{-1}, M_k.



(a) (b) (c) (d) (e)

Figure 2.6: (a) the $N + 6$ control vertices around the extraordinary face f that determine $\phi_i : \Omega \to \mathbb{R}$ (b) the $N + 6$ control vertices at the next subdivision level that determine $\phi_i|_{\frac{1}{2}\Omega}$ (c)-(e) the 12 control vertices at the next subdivision level that determine $\phi_i|_{\Omega_k^1}$, $k = 1, 2, 3$

3. Computation of $W(\mathcal{V})$, $A(\mathcal{V})$, $V(\mathcal{V})$, $M(\mathcal{V})$ and their gradients

To summarize the previous section, a regular patch of a Loop subdivision surface is parametrized by a single degree 4 polynomial on the reference triangle Ω for each of the three spatial components, whereas an irregular patch admits a more complicated parametrization over Ω. In either case, an efficient algorithm exists for evaluating the parameterization and its derivatives at arbitrary parameter values $(v, w) \in \Omega$. Armed with such evaluation algorithms, we now see how the various functionals and their gradient vectors in (2.1) can be computed.

3.1 Formulas for $A(\mathcal{V})$ and $\nabla A(\mathcal{V})$

We first discuss how to compute the area $A(\mathcal{V})$ of a Loop surface and the gradient $\nabla A(\mathcal{V})$ of A w.r.t. $\mathcal{V}$. For each regular face f in a control mesh, we write $A_6(\mathbf{c}_f)$ as the area of the surface patch associated to f; and we view A_6 as a (real-valued) function of the variables in the array $\mathbf{c}_f$. Similarly, we write $A_N(\mathbf{c}_f)$ as the area of the surface patch associated to an irregular face f with an extraordinary vertex of valence $N \neq 6$; in this case, we also write $\mathrm{val}(f) := N$. So

$$A(\mathcal{V}) = \underbrace{\sum_{f \text{ regular}} A_6(\mathbf{c}_f)}_{\text{area of regular patches}} + \underbrace{\sum_{N \neq 6} \sum_{f:\, \mathrm{val}(f)=N} A_N(\mathbf{c}_f)}_{\text{area of irregular patches}}. \tag{3.1}$$

Recall from (2.5) and (2.7), the parametrization $\mathbf{s}_f$ can be written as

$$\mathbf{s}_f = \mathbf{c}_f^T \mathbf{b}^N, \text{ where } N = \mathrm{val}(f). \tag{3.2}$$

Note that $\mathbf{c}_f$ is *linearly* related to $\mathcal{V}$, whereas the basis functions $\mathbf{b}^N$ are *independent*

of $\mathcal{V}$. These features allow for an accurate and efficient computation of $A(\mathcal{V})$ and $\nabla A(\mathcal{V})$, as we shall now see.

We define the map P_f by

$$P_f \mathcal{V} = \mathbf{c}_f \tag{3.3}$$

which picks out the local control data around the face f from the global control data $\mathcal{V}$.

For convenience, drop the subscript and write $\mathbf{s}$ instead of $\mathbf{s}_f$. Also, we write $\mathbf{s}_1$, $\mathbf{s}_2$, $\mathbf{s}_3$ for the components of $\mathbf{s}$ and $\mathbf{s}_{i,u}$, $\mathbf{s}_{i,v}$ for their partial derivatives.

Note that

$$\begin{aligned} A_N(\mathbf{c}_f) &= \iint_\Omega \|\mathbf{n}(v,w)\|\, dv\, dw, \text{ where} \\ \mathbf{n} &= \frac{\partial \mathbf{s}}{\partial v} \times \frac{\partial \mathbf{s}}{\partial w} = \left[\mathbf{s}_{2,v}\mathbf{s}_{3,w} - \mathbf{s}_{2,w}\mathbf{s}_{3,v},\ \ \mathbf{s}_{3,v}\mathbf{s}_{1,w} - \mathbf{s}_{3,w}\mathbf{s}_{1,v},\ \ \mathbf{s}_{1,v}\mathbf{s}_{2,w} - \mathbf{s}_{1,w}\mathbf{s}_{2,v}\right]. \end{aligned} \tag{3.4}$$

Again, we drop the subscript f and write $\mathbf{c}_{\cdot 1}$, $\mathbf{c}_{\cdot 2}$, $\mathbf{c}_{\cdot 3}$ to refer to the columns of $\mathbf{c}_f$.

When f is a regular face, $\mathbf{s}_i = \mathbf{c}_{\cdot i}^T \mathbf{b}$, so

$$\mathbf{s}_{i,v} = \mathbf{c}_{\cdot i}^T \mathbf{b}_v, \ \ \mathbf{s}_{i,w} = \mathbf{c}_{\cdot i}^T \mathbf{b}_w, \ \ i = 1,2,3. \tag{3.5}$$

The gradient of $\mathbf{n}_1$ with respect to $\mathbf{c}_f$, organized as a 12×3 array (i.e. same

dimension as $\mathbf{c}_f$), can be expressed as:[1]

$$\begin{aligned}\nabla_{\mathbf{c}_f}\mathbf{n}_1 &= \nabla_{\mathbf{c}_f}(\mathbf{s}_{2,v}\mathbf{s}_{3,w} - \mathbf{s}_{2,w}\mathbf{s}_{3,v}) \\ &= \big[\mathbf{0},\ \ \mathbf{b}_v,\ \ \mathbf{0}\big]\mathbf{s}_{3,w} + \mathbf{s}_{2,v}\big[\mathbf{0},\ \ \mathbf{0},\ \ \mathbf{b}_w\big] - \big[\mathbf{0},\ \ \mathbf{b}_w,\ \ \mathbf{0}\big]\mathbf{s}_{3,v} - \mathbf{s}_{2,w}\big[\mathbf{0},\ \ \mathbf{0},\ \ \mathbf{b}_v\big] \\ &= \big[\mathbf{0},\ \ \mathbf{s}_{3,w}\mathbf{b}_v - \mathbf{s}_{3,v}\mathbf{b}_w,\ \ \mathbf{s}_{2,v}\mathbf{b}_w - \mathbf{s}_{2,w}\mathbf{b}_v\big].\end{aligned} \tag{3.6}$$

Similarly,

$$\begin{aligned}\nabla_{\mathbf{c}_f}\mathbf{n}_2 &= \big[-\mathbf{s}_{3,w}\mathbf{b}_v + \mathbf{s}_{3,v}\mathbf{b}_w,\ \ \mathbf{0},\ \ -\mathbf{s}_{1,v}\mathbf{b}_w + \mathbf{s}_{1,w}\mathbf{b}_v\big], \\ \nabla_{\mathbf{c}_f}\mathbf{n}_3 &= \big[\mathbf{s}_{2,w}\mathbf{b}_v - \mathbf{s}_{2,v}\mathbf{b}_w,\ \ \mathbf{s}_{1,v}\mathbf{b}_w - \mathbf{s}_{1,w}\mathbf{b}_v\ \ \mathbf{0}\big].\end{aligned} \tag{3.7}$$

Next, we have

$$\begin{aligned}\|\mathbf{n}\| &= \sqrt{\langle\mathbf{n},\mathbf{n}\rangle}, \\ \nabla_{\mathbf{c}_f}\|\mathbf{n}\| &= \nabla_{\mathbf{c}_f}\langle\mathbf{n},\mathbf{n}\rangle^{1/2} = \frac{1}{2\langle\mathbf{n},\mathbf{n}\rangle^{1/2}}\nabla_{\mathbf{c}_f}\langle\mathbf{n},\mathbf{n}\rangle \\ &= \frac{1}{\langle\mathbf{n},\mathbf{n}\rangle^{1/2}}\left(\mathbf{n}_1\nabla_{\mathbf{c}_f}\mathbf{n}_1 + \mathbf{n}_2\nabla_{\mathbf{c}_f}\mathbf{n}_2 + \mathbf{n}_3\nabla_{\mathbf{c}_f}\mathbf{n}_3\right).\end{aligned} \tag{3.8}$$

Therefore, the local area functional A_N and its gradient

$$\nabla A_N(\mathbf{c}_f) = \iint_\Omega \nabla_{\mathbf{c}_f}\|\mathbf{n}(v,w)\|\,dv\,dw, \tag{3.9}$$

can be computed based on the control data $\mathbf{c}_f = [\mathbf{c}_{\cdot 1}, \mathbf{c}_{\cdot 2}, \mathbf{c}_{\cdot 3}]$ and the basis function $\mathbf{b}$ via (3.5)-(3.8). Together with (3.3) and the chain rule, we can compute the total

[1]Here and below, we have to deal with a number of scalar quantities S that vary with **both** the local control data $\mathbf{c}_f$ and parameter values (v,w) (e.g. $\|\mathbf{n}\|$, $\mathbf{n}_i$, E, F, G, e, f, g, etc..) In order to avoid confusion, we use the notation '$\nabla_{\mathbf{c}_f}S$' to denote the gradient vector of S viewed as a function of $\mathbf{c}_f$; the gradient 'vector' is structured as an array of the same size as $\mathbf{c}_f$. Likewise, the gradient 'vectors' $\nabla W(\mathcal{V})$, $\nabla A(\mathcal{V})$, $\nabla V(\mathcal{V})$, $\nabla M(\mathcal{V})$ are structured as $\#V \times 3$ arrays, i.e. the same dimensions as $\mathcal{V}$.

area and its gradient with respect to $\mathcal{V}$ by

$$A(\mathcal{V}) = \sum_N \sum_{f:\,\mathrm{val}(f)=N} A_N(P_f\mathcal{V}), \quad \nabla A(\mathcal{V}) = \sum_N \sum_{f:\,\mathrm{val}(f)=N} P_f^T \nabla A_N(\mathbf{c}_f). \tag{3.10}$$

3.2 Formulas for $W(\mathcal{V})$, $V(\mathcal{V})$, $M(\mathcal{V})$, and their gradients

Similar to $A(\mathcal{V})$, we aim to express the other three functionals in the C-H model in terms of $\mathbf{c}_f$ and the basis functions $\mathbf{b}$ and Φ. For W and M, we need an expression for the mean curvature. Recall that

$$E = \langle \mathbf{s}_{f,v}, \mathbf{s}_{f,v} \rangle,\ F = \langle \mathbf{s}_{f,v}, \mathbf{s}_{f,w} \rangle,\ G = \langle \mathbf{s}_{f,w}, \mathbf{s}_{f,w} \rangle$$

represent the first fundamental form of the surface $\mathbf{s}_f$, whereas

$$e = \langle \mathbf{s}_{f,vv}, \mathbf{n} \rangle / \|\mathbf{n}\|,\ f = \langle \mathbf{s}_{f,vw}, \mathbf{n} \rangle / \|\mathbf{n}\|,\ g = \langle \mathbf{s}_{f,ww}, \mathbf{n} \rangle / \|\mathbf{n}\|$$

represent the second fundamental form. The mean curvature can be expressed as

$$H = \frac{eG - 2fF + gE}{2(EG - F^2)}. \tag{3.11}$$

Therefore

$$M = \iint H\, dA = \sum_{f\in\mathcal{F}} \iint_\Omega \frac{eG - 2fF + gE}{2(EG - F^2)} \|\mathbf{n}\|\, dv\, dw = \sum_{f\in\mathcal{F}} \iint_\Omega \frac{\bar{e}G - 2\bar{f}F + \bar{g}E}{2(EG - F^2)}\, dv\, dw, \tag{3.12}$$

where $\bar{e} := \langle \mathbf{s}_{f,vv}, \mathbf{n} \rangle$, $\bar{f} := \langle \mathbf{s}_{f,vv}, \mathbf{n} \rangle$, $\bar{g} := \langle \mathbf{s}_{f,vv}, \mathbf{n} \rangle$. Similarly,

$$W = \iint H^2\, dA = \sum_{f\in\mathcal{F}} \iint_\Omega \left[\frac{\bar{e}G - 2\bar{f}F + \bar{g}E}{2(EG - F^2)} \right]^2 \frac{1}{\|\mathbf{n}\|}\, dv\, dw. \tag{3.13}$$

Remark. When f is an irregular face, the mean curvature can potentially blows up when approaching the extraordinary vertex, however it is proved in [31] that the corresponding integrals (called M_N and W_N below) are always finite.

By the divergence theorem, with the choice of the vector field $\vec{X}(x,y,z) = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$, the volume enclosed by a surface can be expressed as a surface integral:

$$V = \frac{1}{3}\iiint \mathrm{div}\vec{X}\, dx\, dy\, dz = \frac{1}{3}\iint \vec{X}\cdot\mathbf{n}/\|\mathbf{n}\|\, dA = \frac{1}{3}\sum_{f\in\mathcal{F}}\iint_{\Omega}\langle \mathbf{s}_f, \mathbf{n}\rangle\, dv\, dw. \tag{3.14}$$

By (3.2), we can express the rightmost integral in (3.12)-(3.14) in terms of $\mathbf{c}_f$ and $\mathbf{b}^N$ when where f ranges over all faces and N ranges over all valences existing in the control mesh; we denote the integral by $M_N(\mathbf{c}_f)$, $W_N(\mathbf{c}_f)$ and $V_N(\mathbf{c}_f)$, respectively.

We explain how to compute the gradients of $M_N(\mathbf{c}_f)$, $W_N(\mathbf{c}_f)$ and $V_N(\mathbf{c}_f)$.

The gradients of E, F, G, $\bar{e}$, $\bar{f}$, $\bar{g}$ can be computed as follows

$$\begin{aligned}
&\mathbf{s}_i = \mathbf{c}_{\cdot i}^T\mathbf{b},\ \mathbf{s}_{i,v} = \mathbf{c}_{\cdot i}^T\mathbf{b}_v,\ \mathbf{s}_{i,v} = \mathbf{c}_{\cdot i}^T\mathbf{b}_w,\ \mathbf{s}_{i,vv} = \mathbf{c}_{\cdot i}^T\mathbf{b}_{vv},\ \mathbf{s}_{i,vw} = \mathbf{c}_{\cdot i}^T\mathbf{b}_{vw},\ \mathbf{s}_{i,ww} = \mathbf{c}_{\cdot i}^T\mathbf{b}_{ww},\\
&\nabla_{\mathbf{c}_f}E = \sum_{i=1,2,3}\nabla_{\mathbf{c}_f}(\mathbf{c}_{\cdot i}^T\mathbf{b}_v)^2 = 2\big[\mathbf{s}_{1,v}\mathbf{b}_v,\ \mathbf{s}_{2,v}\mathbf{b}_v,\ \mathbf{s}_{3,v}\mathbf{b}_v\big],\\
&\nabla_{\mathbf{c}_f}F = \sum_{i=1,2,3}\nabla_{\mathbf{c}_f}(\mathbf{c}_{\cdot i}^T\mathbf{b}_v)(\mathbf{c}_{\cdot i}^T\mathbf{b}_w) = \big[\mathbf{s}_{1,v}\mathbf{b}_w + \mathbf{s}_{1,w}\mathbf{b}_v,\ \mathbf{s}_{2,v}\mathbf{b}_w + \mathbf{s}_{2,w}\mathbf{b}_v,\ \mathbf{s}_{3,v}\mathbf{b}_w + \mathbf{s}_{3,w}\mathbf{b}_v\big],\\
&\nabla_{\mathbf{c}_f}G = \sum_{i=1,2,3}\nabla_{\mathbf{c}_f}(\mathbf{c}_{\cdot i}^T\mathbf{b}_w)^2 = 2\big[\mathbf{s}_{1,w}\mathbf{b}_w,\ \mathbf{s}_{2,w}\mathbf{b}_w,\ \mathbf{s}_{3,w}\mathbf{b}_w\big],\\
&\nabla_{\mathbf{c}_f}\bar{e} = \sum_i \nabla_{\mathbf{c}_f}(\mathbf{c}_{\cdot i}^T\mathbf{b}_{vv})\,\mathbf{n}_i = \big[\mathbf{s}_{1,vv}\nabla_{\mathbf{c}_f}\mathbf{n}_1,\ \mathbf{s}_{2,vv}\nabla_{\mathbf{c}_f}\mathbf{n}_2,\ \mathbf{s}_{3,vv}\nabla_{\mathbf{c}_f}\mathbf{n}_3\big] + \big[\mathbf{n}_1\mathbf{b}_{vv},\ \mathbf{n}_2\mathbf{b}_{vv},\ \mathbf{n}_3\mathbf{b}_{vv}\big],\\
&\nabla_{\mathbf{c}_f}\bar{f} = \big[\mathbf{s}_{1,vw}\nabla_{\mathbf{c}_f}\mathbf{n}_1,\ \mathbf{s}_{2,vw}\nabla_{\mathbf{c}_f}\mathbf{n}_2,\ \mathbf{s}_{3,vw}\nabla_{\mathbf{c}_f}\mathbf{n}_3\big] + \big[\mathbf{n}_1\mathbf{b}_{vw},\ \mathbf{n}_2\mathbf{b}_{vw},\ \mathbf{n}_3\mathbf{b}_{vw}\big],\\
&\nabla_{\mathbf{c}_f}\bar{g} = \big[\mathbf{s}_{1,ww}\nabla_{\mathbf{c}_f}\mathbf{n}_1,\ \mathbf{s}_{2,ww}\nabla_{\mathbf{c}_f}\mathbf{n}_2,\ \mathbf{s}_{3,ww}\nabla_{\mathbf{c}_f}\mathbf{n}_3\big] + \big[\mathbf{n}_1\mathbf{b}_{ww},\ \mathbf{n}_2\mathbf{b}_{ww},\ \mathbf{n}_3\mathbf{b}_{ww}\big].
\end{aligned} \tag{3.15}$$

By (3.12),

$$\nabla M_N = \nabla_{\mathbf{c}_f} \iint_\Omega \underbrace{\frac{\bar{e}G - 2\bar{f}F + \bar{g}E}{2(EG - F^2)}}_{=:\bar{H}} \, dv\, dw = \iint_\Omega \nabla_{\mathbf{c}_f} \bar{H} \, dv\, dw. \tag{3.16}$$

Thanks to (3.15), the integrand $\nabla_{\mathbf{c}_f} \bar{H}$ can be computed by the product and quotient rules. After ∇M_N is computed, ∇W_N can then be computed by

$$\begin{aligned} \nabla W_N \quad &= \nabla_{\mathbf{c}_f} \textstyle\iint_\Omega \left[\frac{\bar{e}G - 2\bar{f}F + \bar{g}E}{2(EG-F^2)}\right]^2 \frac{1}{\|\mathbf{n}\|} \, dv\, dw \\ &= \textstyle\iint_\Omega \frac{2\bar{H}}{\|\mathbf{n}\|} \nabla_{\mathbf{c}_f} \bar{H} - \frac{\bar{H}^2}{\|\mathbf{n}\|^2} \nabla_{\mathbf{c}_f} \|\mathbf{n}\| \, dv\, dw. \end{aligned} \tag{3.17}$$

Note that every term in the integrand of (3.17) was computed previously.

Finally, for V_N we have

$$\begin{aligned} \nabla V_N &= \nabla_{\mathbf{c}_f} \iint_\Omega \langle \mathbf{s}_f, \mathbf{n} \rangle \, dv\, dw = \iint_\Omega \sum_i \nabla_{\mathbf{c}_f} (\mathbf{c}_{:i}^T \mathbf{b}) \mathbf{n}_i \, dv\, dw \\ &= \iint_\Omega \left\{ \left[\mathbf{n}_1 \mathbf{b},\ \mathbf{n}_2 \mathbf{b},\ \mathbf{n}_3 \mathbf{b}\right] + \mathbf{s}_1 \nabla_{\mathbf{c}_f} \mathbf{n}_1 + \mathbf{s}_2 \nabla_{\mathbf{c}_f} \mathbf{n}_2 + \mathbf{s}_3 \nabla_{\mathbf{c}_f} \mathbf{n}_3 \right\} dv\, dw. \end{aligned} \tag{3.18}$$

With all the local functionals and their gradients w.r.t. to the local control data computed, the global functionals and their gradients w.r.t. to the global control data can be computed exactly as in (3.10):

$$M(\mathcal{V}) = \sum_N \sum_{f:\, \mathrm{val}(f)=N} M_N(P_f \mathcal{V}), \quad \nabla M(\mathcal{V}) = \sum_N \sum_{f:\, \mathrm{val}(f)=N} P_f^T \nabla M_N(\mathbf{c}_f), \tag{3.19}$$

and similarly for $W(\mathcal{V})$, $\nabla W(\mathcal{V})$, $V(\mathcal{V})$, $\nabla V(\mathcal{V})$.

3.3 Implementation details

In the actual numerical computation of $A(\mathcal{V})$ and $\nabla A(\mathcal{V})$ based on (3.10), we use a symmetric 7-point Gauss quadrature rule on a triangle, with accuracy order

5 (see for example [13]), to approximate the integrals in $A_N(\mathbf{c}_f)$, $V_N(\mathbf{c}_f)$, $M_N(\mathbf{c}_f)$, $W_N(\mathbf{c}_f)$, and their gradients (recall (3.9), (3.16)-(3.18)). The approximation is done using a uniform grid of size $1/n$; see Figure 3.1. Typically we use $n = 8$ or 16 at coarse subdivision levels, and $n = 1$ or 2 at fine levels. We choose n to be a dyadic integer in order to take advantage of the subdivision structure of $\mathbf{s}_f$ when $N \neq 6$; see Figure 3.1(a) for $n = 2$. A key implementation detail is that quantities independent

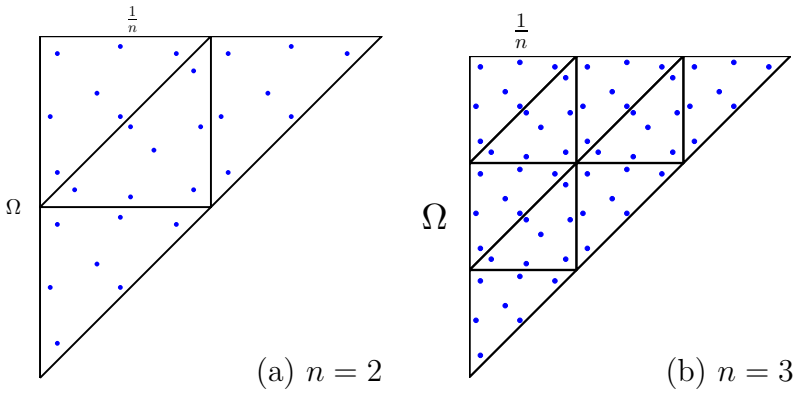


Figure 3.1: A uniform grid on Ω with n^2 sub-triangles, and the 7 quadrature points on each sub-triangle.

of $\mathcal{V}$ are precomputed before entering the optimization loop. These quantities include $\mathbf{b}^6(u, v)$ and $\mathbf{b}^N(u, v) = V_N^{-T}\Phi_N(u, v)$ (for only those extraordinary valences N that show up in $\mathcal{F}$) and their derivatives at the quadrature points (Figure 3.1), evaluated using Stam's algorithm (Sections 2.3-2.4.) A separate pre-processing step computes the maps P_f for each face f in $\mathcal{F}$, as P_f depends only on the connectivity information in $\mathcal{F}$. Note: One should not store P_f as a $(N+6) \times \#V$ matrix as suggested by (3.3). Instead, it suffices to store the information in P_f by a list of $N + 6$ integer indices in $\{1, \ldots, \#V\}$ which keeps track of the indices of $\mathbf{c}_f$ in the global vertex list $\mathcal{V}$; we

denote this list of vertex indices by

$$\mathrm{VFL}(f).$$

These preprocessing steps speed up our solver significantly already in a sequential implementation; Figure 3.2 illustrates the basic structure of our solver.

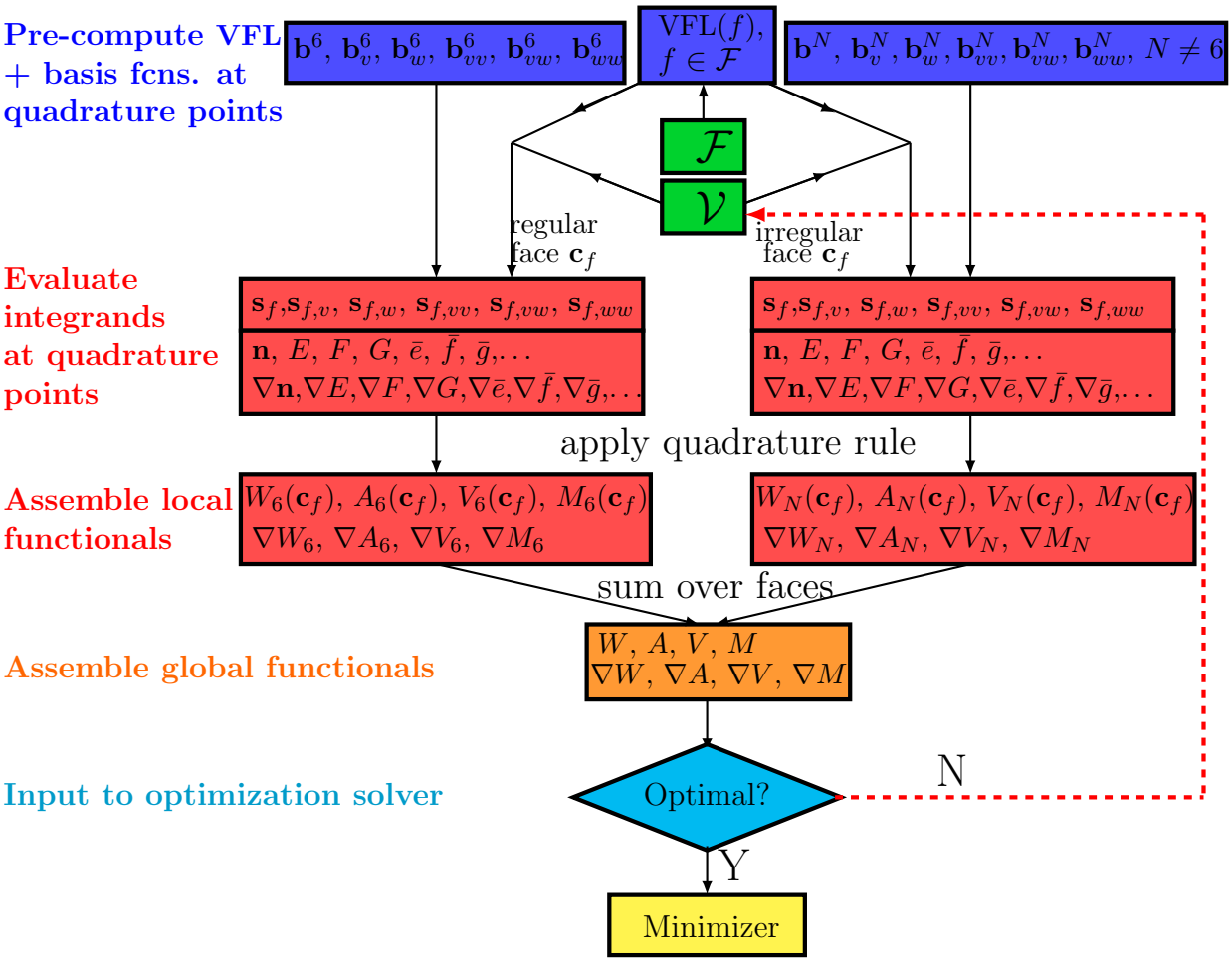


Figure 3.2: Flow chart of the monoscale version of our solver

3.4 Parallel implementation

3.4.1 Overall architecture

Since the computations of $W(\mathcal{V})$, $A(\mathcal{V})$, $V(\mathcal{V})$, $M(\mathcal{V})$ and their gradients are the bottleneck of the whole solver, we seek to parallelize them in a GPU, and leave the nonlinear optimization solver running in the CPU. Recall that each iteration of the optimization process updates the values in $\mathcal{V}$ (see Figure 3.2), so the only data that needs to be communicated between the CPU and the GPU at each iteration are:

- From CPU to GPU: $\mathcal{V}$
- From GPU to CPU: $W(\mathcal{V})$, $A(\mathcal{V})$, $V(\mathcal{V})$, $M(\mathcal{V})$, $\nabla W(\mathcal{V})$, $\nabla A(\mathcal{V})$, $\nabla V(\mathcal{V})$, $\nabla M(\mathcal{V})$;

see Figure 3.3. The communication cost is in the order of $O(\#\text{vertices})$. In our experience, it is negligible compared to the time spent on the computation of the objective and constraint functionals and their first variations. We now discuss our CUDA implementation of $W(\mathcal{V})$, $A(\mathcal{V})$, $V(\mathcal{V})$ and $M(\mathcal{V})$ and their gradients.

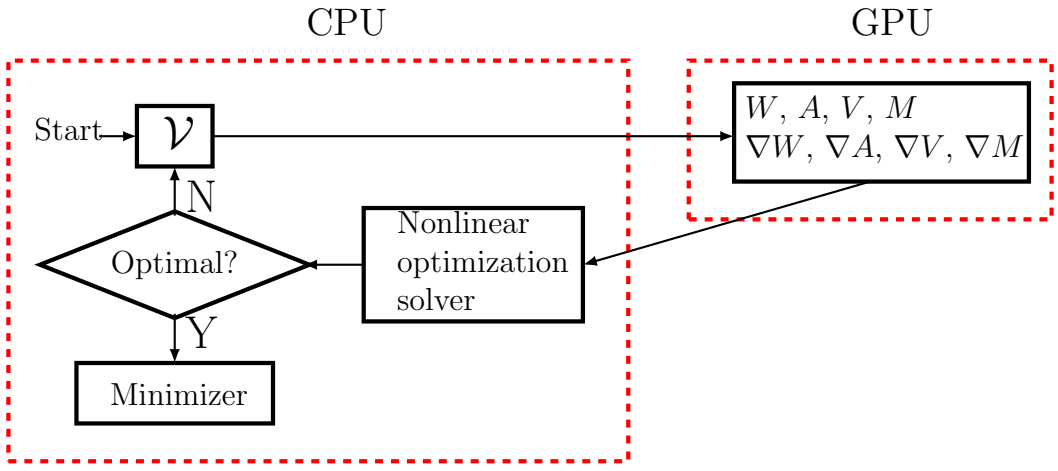


Figure 3.3: Architecture of our GPU enchanced solver

3.4.2 Parallel computation of $W(\mathcal{V})$, $A(\mathcal{V})$, $V(\mathcal{V})$ and $M(\mathcal{V})$

Recall from Figure 3.2 that the computation of the functionals $W(\mathcal{V})$, $A(\mathcal{V})$, $V(\mathcal{V})$ and $M(\mathcal{V})$ require the evaluation of the surfaces patches $\mathbf{s}_f(u, v)$ and their first and second partial derivatives. This otherwise costly computation can be carried out in parallel for all faces $f \in \mathcal{F}$ and all quadrature points (u, v) (recall Figure 3.1). In our CUDA implementation, we use a two-dimensional grid of thread blocks, with one dimension representing the different faces and the other dimension representing different quadrature points. The $\#F$ 'local functionals' can then be computed in parallel. Subsequently, these local functionals are summed to the resulting values of $W(\mathcal{V})$, $A(\mathcal{V})$, $V(\mathcal{V})$ and $M(\mathcal{V})$.

3.4.3 Parallel computation of $\nabla W(\mathcal{V})$, $\nabla A(\mathcal{V})$, $\nabla V(\mathcal{V})$ and $\nabla M(\mathcal{V})$

The parallel computation of the gradient vectors is trickier. Consider, for instance, our sequential computation for $\nabla A(\mathcal{V})$ based on (3.10); recall Section 3.3. In it, we loop through all the faces $f \in \mathcal{F}$ and add each row of the (short, local) gradient vector $\nabla A_N(\mathbf{c}_f)$ to the appropriate row of the (long, global) gradient vector $\nabla A(\mathcal{V})$. This process is facilitated by the pre-computed data $\mathrm{VFL}(f)$. If we attempt to parallelize the computation over the faces, two different faces may contribute simultaneously to the same row of $\nabla A(\mathcal{V})$, thus causing a potential *race condition.* Race condition can be avoided by *atomic operations* in CUDA. In the case at hand, we can use the CUDA function `atomicAdd()`.

Instead of parallelizing over the faces, we can parallelize over the vertices, i.e. we compute the rows of $\nabla A(\mathcal{V})$ in parallel. Such a vertex-based parallel algorithm does not require atomic operations; instead it requires us to re-express the second identity in (3.10). Below we describe our parallel algorithm for $\nabla A(\mathcal{V})$, that for the other gradient vectors are analogous.

Recall from Section 3.1 that our gradient vector $\nabla A(\mathcal{V})$ is structured as a $\#V \times 3$ array (i.e. the same data-structure for the vertex list $\mathcal{V}$), with the i-th row being the rate of change of the area functional w.r.t. the coordinates of the i-th vertex, i.e.

$$\left[\frac{\partial A}{\partial \mathcal{V}_{i,x}}, \frac{\partial A}{\partial \mathcal{V}_{i,y}}, \frac{\partial A}{\partial \mathcal{V}_{i,z}}\right] =: \nabla A(\mathcal{V})_i. \tag{3.20}$$

Note that, for each vertex index i, this row vector is only dependent on the gradients of the local surface area functionals $\nabla A_N(\mathbf{c}_f)$ for a small number of faces. Therefore, we pre-compute the list of faces in $\mathcal{F}$, denoted here by $\mathrm{FVL}(i)$, that contributes to $\nabla A(\mathcal{V})_i$. We then notice that the second identity in (3.10) can be rewritten as:

$$\nabla A(\mathcal{V})_i = \sum_{f \in \mathrm{FVL}(i)} \left[\frac{\partial A_{\mathrm{val}(f)}}{\partial \mathcal{V}_{i,x}}(\mathbf{c}_f), \frac{\partial A_{\mathrm{val}(f)}}{\partial \mathcal{V}_{i,y}}(\mathbf{c}_f), \frac{\partial A_{\mathrm{val}(f)}}{\partial \mathcal{V}_{i,z}}(\mathbf{c}_f)\right] \tag{3.21a}$$

$$= \sum_{f \in \mathrm{FVL}(i)} \left(\nabla A_{\mathrm{val}(f)}(\mathbf{c}_f)\right)_{\mathrm{r}(i,f)}. \tag{3.21b}$$

Note that $[\mathcal{V}_{i,x}, \mathcal{V}_{i,y}, \mathcal{V}_{i,z}]$ is one of the rows of $\mathbf{c}_f$, and $[\frac{\partial A_{\mathrm{val}(f)}}{\partial \mathcal{V}_{i,x}}(\mathbf{c}_f), \frac{\partial A_{\mathrm{val}(f)}}{\partial \mathcal{V}_{i,y}}(\mathbf{c}_f), \frac{\partial A_{\mathrm{val}(f)}}{\partial \mathcal{V}_{i,z}}(\mathbf{c}_f)]$ is the corresponding row of $\nabla A_{\mathrm{val}(f)}(\mathbf{c}_f)$; we denote this row number by $\mathrm{r}(i,f)$ in (3.21b). This formula allows us to efficiently compute the rows of the global gradient vector $\nabla A(\mathcal{V})$ in parallel from the gradients of the local area functionals.

The auxiliary data structures $\mathrm{FVL}(i)$, $i = 1, \ldots, \#V$, $\mathrm{r}(i,f)$, $f \in \mathrm{FVL}(i)$, that facilitate the parallel algorithm are only dependent on the connectivity information in $\mathcal{F}$, meaning that they only need to be computed once and stored in GPU memory before the optimization loop begins.

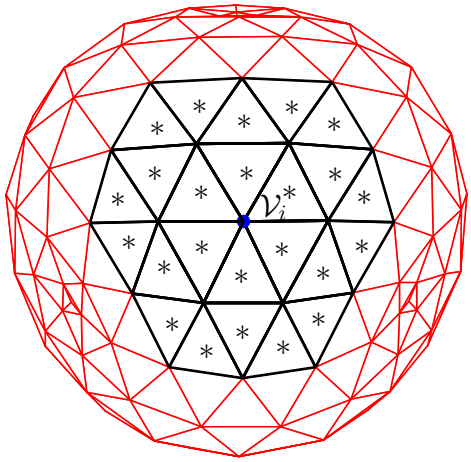


Figure 3.4: FVL(i): for the Loop scheme it is the 2-ring of faces around vertex i

4. Other Versions of the Solver

4.1 Multiscale version

The multiscale and local nature of subdivision surfaces allow us to first optimize (quickly) at a coarse scale, then subdivide to gain more degrees of freedom, optimize again at the finer scale, then subdivide again, and so on. We found such a multi-scale optimization approach extremely efficient, especially when combined with the parallel implementation; in fact it is indispensable when one desires a considerably accurate solution in the difficult regimes of the problems, e.g. when the reduced volume constraint value v_0 approaches zero in the Canham problem, see Section 5.1.

4.2 Genus 0 and 1 surface of revolution version

In many cases, the solution of the Canham or Helfrich problem appears to be a surface of revolution. In verifying such a conjecture, it is helpful to solve the problem by constraining S in (1.1) to be a genus 0 or 1 surface of revolution. (Note: By Hurwitz's automorphisms theorem, a surface of genus $g > 1$ cannot be a surface of revolution.) The Loop scheme can approximate but cannot exactly produce surfaces of revolution. Using cubic B-spline subdivision curves to approximate the profile curve, we can solve the Canham and Helfrich problems in this further constrained case. In Section 5, we shall use the general solver together with the surface of revolution solver to explore when the surface of revolution symmetry persists or breakdowns in the problems.

In the genus 1 case, the profile curve is a closed curve, and we use a periodic control polygon to model the profile curve; see the right panel of Figure 4.1. In the genus 0, we use an open curve which contacts the axis of revolution orthogonally

at the two ends, so that the resulted surface of revolution has a finite Willmore energy. In order to generate such a cubic B-Spline curve, we use a control polygon $[C_{-1}, C_0, C_1, \ldots, C_{N-1}, C_N]$, $C_i \in \mathbb{R}^2$, with only the central N vertices to be free (i.e. treated as degrees of freedom in the optimization), and impose the boundary condition:

$$C_{-1} = (C_{1,x}, -4C_{0,y} - C_{1,y}), \quad C_N = (C_{N-2,x}, -4C_{N-1,y} - C_{N-2,y}).$$

Here we assume the horizontal x-axis is the axis or revolution. See the left panel of Figure 4.1.

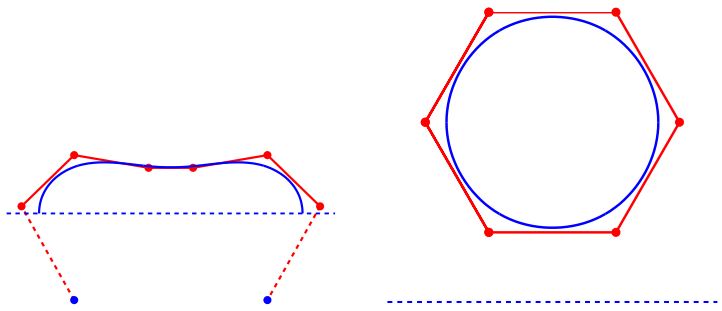


Figure 4.1: Uniform cubic B-Spline curve and control polygon for a genus 0 (left) and genus 1 (right) surface of revolution. There are $N = 6$ free control vertices in each case.

4.3 Special higher order genus 0 (and 1) version

Since the spherical topology is arguably the most important topology in cell biology, couldn't we simply use the most classical approximation, namely *spherical harmonics* for this very case? Indeed, our monoscale method can be applied by replacing the local subdivision basis functions with the global spherical harmonics up

to a certain degree, i.e. $\{Y_\ell^m : \ell = 0, 1, \ldots, \ell_0, -\ell \leqslant m \leqslant \ell\}$. As there is no looping and summing over triangle faces, the implementation of such a spherical harmonics approach is relatively simple. However, because the basis functions are global, it is fundamentally difficult to find a good set of Fourier coefficients to initiate the optimization loop, even when we have a reasonable idea of how the surface should look like. Unlike the control vertices in a subdivision scheme, the Fourier coefficients do not control the shape of a surface in an intuitive way. The multiscale idea does not help much either: assume that we manage to solve the optimization problem very accurately up to degree ℓ_0, and desire to expand the set of spherical harmonics to degree $\ell_0 + 1$, we may naturally set the initial guess to be the Fourier coefficients from the previous optimization for the lower degree harmonics, and zeros for the newly added harmonics. This does not deliver the efficiency one may expect from a multiscale optimization method, again because of the global nature of the harmonics, which causes a slow decay in the Fourier expansions needed to capture a surface with high curvature regions (e.g. the stomatocyte-shaped solutions of the Canham problem when $v_0 \to 0$, see Figure.)

This experience prompted us to use a subdivision schemes specially designed for the genus 0 case [16]. This scheme is based on a sixth-order accuracy subdivision scheme derived from a box-spline in the regular setting,[1] together with a subtle 'miracle' for valence 3 vertices. The scheme is not only more accurate in the regular setting, but also produces flexible C^2 surfaces, i.e. free of the kind of curvature singularities or flat spots at extraordinary vertices that are provably inevitable in the Loop scheme [29, 30].

Surfaces with genus 0 or 1 can be triangulated in such a way that only valence 3 and 6 vertices are involved; see Figure 4.2. (Conversely, it can be shown by Euler's

[1] In comparison, the Loop scheme is based on a fourth-order accurate box-spline.

relation that the closed surfaces that are modelled with only valence 3 and 6 vertices must be of genus 0 or 1.) The special scheme developed in [16] can thus be applied to solve the Canham or Helfrich problem in the genus 0 and 1 cases.

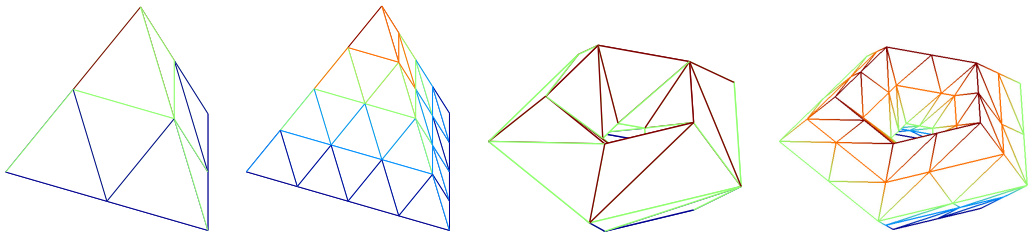


Figure 4.2: Left 2 panels: Genus 0 surfaces with only valence 3 and 6 vertices (subdivided tetrahedra). Right 2 panels: (4×5 and 8×10) regularly triangulated tori, with only valence 6 vertices.

The special scheme exhibits a faster rate of convergence than the Loop scheme for the (highly nonlinear) geometric variational problem, as one may expect from the approximation theory well-studied in the linear functional setting (see, e.g., [14]). See the next section.

5. Computational Results

We report some of the results obtained by our solver of the Canham problem and Helfrich problem. Our choice of computational experiments are motivated by the intertwining interests in biophysics, geometric analysis, and numerical analysis. In particular, we explore phase transition, uniqueness and accuracy of solutions.

5.1 Canham problem

The existence of solution for the Canham problem is studied in [32] for the genus 0 case and in [21] for arbitrary genus. Since we are only interested in the shape of the solution surface up to similarity transformations (a.k.a. homotheties), the solution is characterized by the so-called reduced volume:

$$v_0 := \frac{3V_0}{4\pi}(\frac{A_0}{4\pi})^{-\frac{3}{2}}.$$

By the isoperimetric inequality, $v_0 \in (0, 1]$. [1] (In the papers [32, 21], the authors define the Canham problem by constraining instead the isoperimetric ratio $I := (6\sqrt{\pi})^{1/3}V_0^{1/3}/A_0^{1/2}$; but notice that I is just $v_0^{1/3}$.)

5.1.1 Genus 0 case

The genus 0 case is well-documented in the biophysics literature. See Figure 1.2. Empirical results from the literature suggests that for any $v_0 \in (0, 1]$, the solution is (a) unique (up to similarity) and (b) is a surface of revolution. When $v_0 = 1$, the solution is the round sphere, and when $v_0 \to 0$, the solution approaches a "double

[1] $v_0 = 1$ can only be achieved when it is a round sphere. Thus for $v_0 \in (0, 1)$, for surfaces with genus > 0.

sphere" with a 'neck' that looks like a catenoid (a surface with zero mean curvature). The familiar bi-concave shape of red-blood cells happens when $v_0 \in (0.592, 0.651)$; and phase transitions happen at both end points of this interval.

We provide numerical evidences for the surface of revolution conjecture based on comparing the results from the general solver and the surface of revolution version (see Section 4.2). Table 5.1 shows the numerically computed Willmore energies based on the Loop solver, the surface of revolution solver (see Section 4.2), and the higher order scheme (see Section 4.3). In each case, a multiscale optimization strategy (see Section 4.1) is used. Since the first two solvers for general surfaces cannot produce a solution with a lower Willmore energy than the surface of revolution solver, we are led to believe that the true minimizer is a surface of revolution. Table 5.1 shows the result for $v_0 = 0.72$, similar results are obtained for other values of $v_0 \in (0, 1]$.

solver	Loop	Higher order	Surf. of Revol.
W			19.888089372 (6)
	24.983797204 (34)	19.886338133 (34)	19.882221717 (11)
	19.887306650 (130)	19.881875818 (130)	19.881887245 (21)
	19.882187948(514)	19.881863505 (514)	19.8818652224 (41)

Table 5.1: minimizing W-values for the genus 0 Canham problem with $v_0 = 0.72$ from three different solvers and successively finer control meshes. For the Loop and Higher order solvers, the coarsest control mesh is a twice subdivided tetrahedron; see Figure 4.2. For the surface of revolution solver, the coarsest control polygon has $N = 6$ vertices; see Figure 4.1. Thus numbers in parenthesis are the number of control vertices.

5.1.2 Genus 1 case

The Willmore conjecture (now the Marques-Neves theorem) asserts that the lowest W-energy $\int H^2 dA$ among all genus 1 surfaces is $2\pi^2$ and is attained by the 'Clifford cyclides' (recall footnote 1.) While all the Clifford cyclides are congruent to a Möbius geometer, they are not congruent nor similar in the Euclidean metric; in particular, their reduced volumes are not the same. Using Pappus's centroid theorems, one can verify that the reduced volume of the symmetric Clifford torus is

$$v_{\mathrm{SC}} := \frac{3}{2}(2\pi^2)^{-1/4} \approx 0.7116.$$

It can be shown that among all the Clifford cyclides, the symmetric Clifford torus has the lowest reduced volume. Furthermore, by applying the right sphere inversions one can transform the symmetric Clifford torus into an almost-round sphere with an arbitrarily small handle; this also means that the v-values of Clifford cyclides can get arbitrarily close to 1. To conclude:

- the v-values of the Clifford cyclides span the half-interval $[v_{\mathrm{SC}}, 1)$. By the Marques-Neves theorem, for any $v_0 \in [v_{\mathrm{SC}}, 1)$, the solution of the corresponding genus 1 Canham problem must be a Clifford cyclide. Figure 5.1 shows the solution surfaces obtained by our solver for different values of v_0 in this interval, the numerical results are consistent with the theory.

When $v_0 < v_{\mathrm{SC}}$, the solution of the corresponding Canham problem cannot be a Clifford cyclide anymore. Our solver suggests:

- For $v_0 \in [v^*, v_{\mathrm{SC}}]$, the solution is a surface of revolution with a convex, oval-shaped cross section, where $v^* \approx 0.566$.

- For $v_0 \in (0, v^*]$, the solution is a surface of revolution with a non-convex, sickle-shaped cross section; see Figure 5.2.

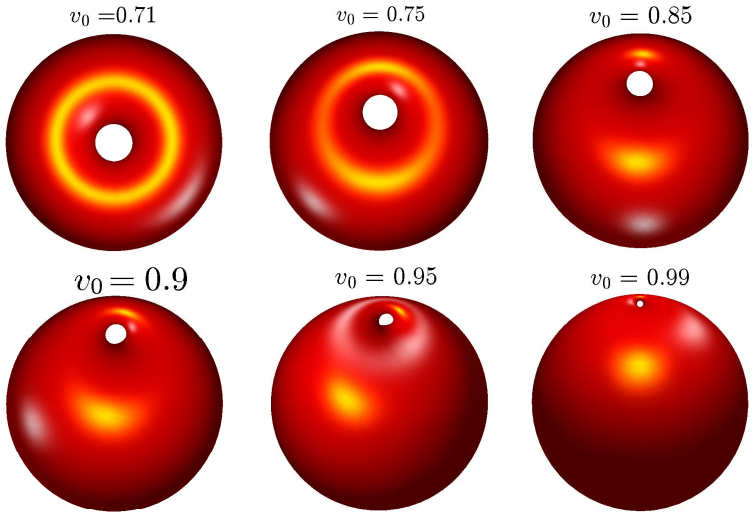


Figure 5.1: Solution surfaces for the genus 1 Canham problem for various $v_0 \in [v_{\mathrm{SC}}, 1)$.

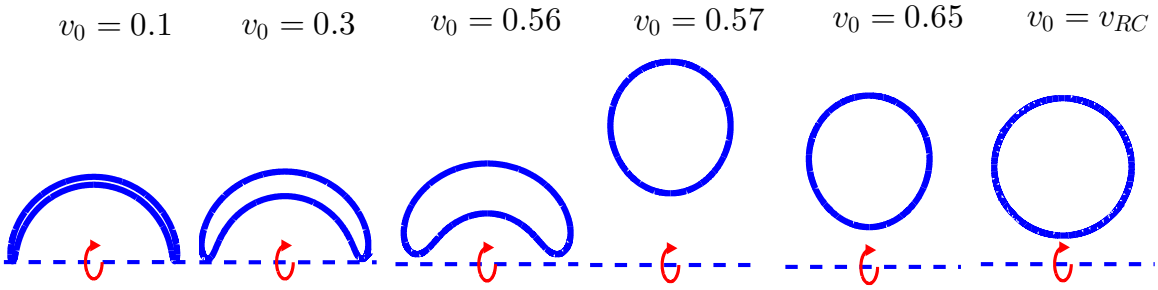


Figure 5.2: Cross-section curves of the solution surfaces for various $v_0 \in (0, v_{\mathrm{SC}}]$

- The problem exhibits a phase transition at v^*. When $v_0 = v^*$, there are two non-equivalent solutions. See Figure 5.3.

Similar to the genus 0 case, we provide numerical evidences for the surface of revolution conjecture in the range $v_0 \in (0, v_{\mathrm{SC}}]$; see Table 5.2.

Figure 5.4(b) plots the computed minimal W-values versus $v_0 \in (0, 1)$. Note that when v_0 is slightly bigger than v^*, a surface of revolution with a sickle-shaped cross-section forms a local (but not global) minimizer; likewise, when v_0 is slightly smaller than v^* or slightly bigger than v_{SC}, a surface of revolution with an oval-shaped

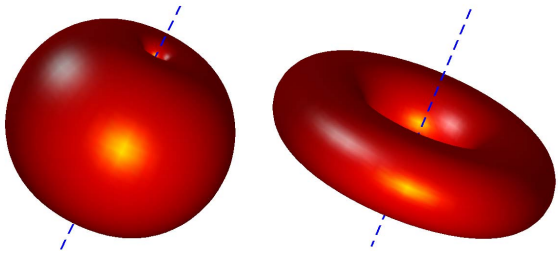


Figure 5.3: The genus 1 Canham problem exhibits a phase transition at $v_0 = v^* \approx 0.566$. In particular, both surfaces above solve the Canham problem at this value of v_0.

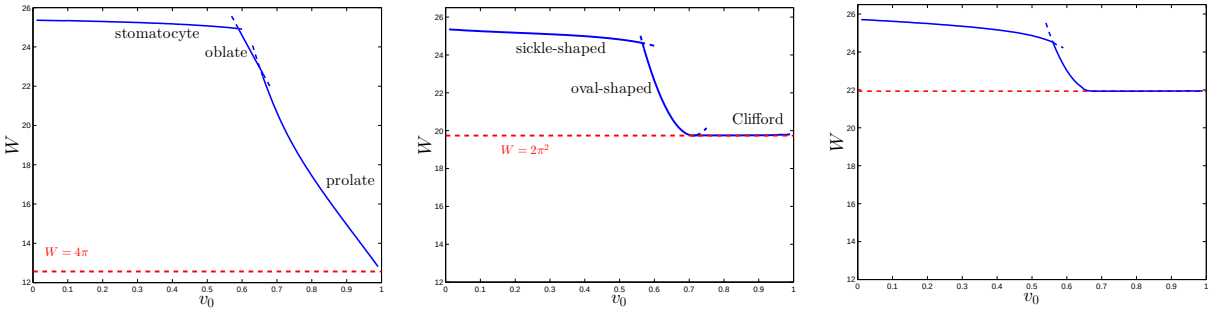
solver	Loop	Higher order	Surf. of Revol.
			24.650383330 (12)
			24.636900846 (24)
W	24.983797204 (50)	24.642763942 (50)	24.635663951 (48)
	24.652875245 (200)	24.635600160 (200)	24.635599721 (96)
	24.636887784 (800)	24.635587574 (800)	24.635592010 (192)
	24.635702627 (3200)		24.635591589 (384)

Table 5.2: minimizing W-values for the genus 1 Canham problem with $v_0 = 0.56$ from three different solvers and successively finer control meshes. For the Loop and Higher order solvers, the coarsest control mesh is a 5×10 regularly triangulated torus; see Figure 4.2. For the surface of revolution solver, the coarsest control polygon has $N = 12$ vertices; see Figure 4.1.

cross-section forms a local minimizer.

5.1.3 Genus 2 case

In [20], it claims the solution to the Helfrich problem is not unique inside a triangular region in the (v_0, m_0) plane and unique outside that region. Thus the solution to the Canham problem is also not unique in the corresponding v_0 range ($[0.65, \infty)$) of the triangular region. Below are the different shapes of minimizers of the Canham problem at $v_0 = 0.66$.



(a) genus 0 (b) genus 1 (c) genus 2

Figure 5.4: Minimal W-values for the genus 0, 1, 2 Canham problem

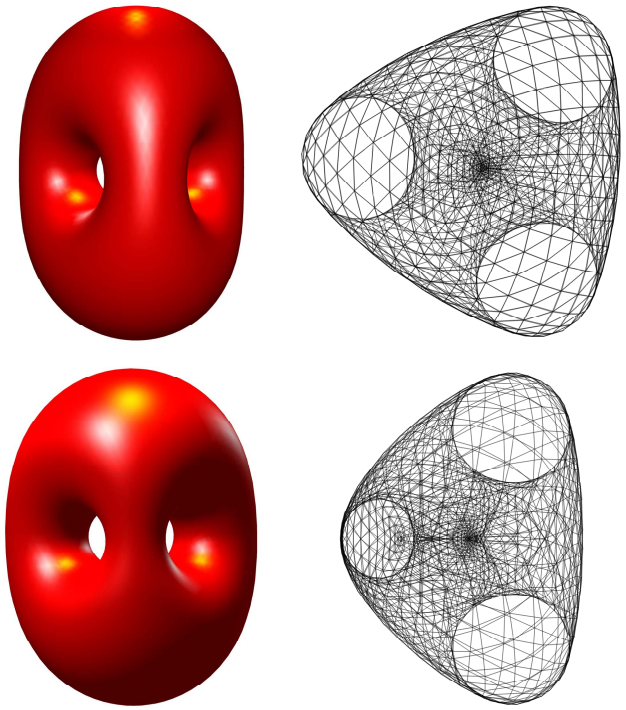


Figure 5.5: Minimizers and their cross sections for genus $= 2$ and $v_0 = 0.66$

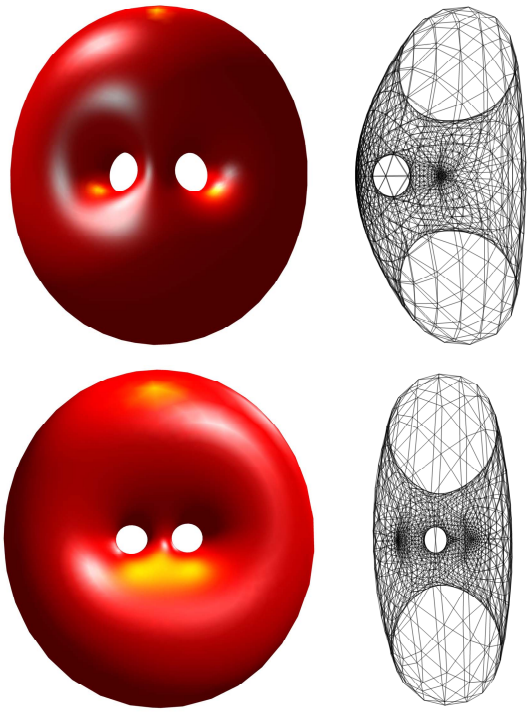


Figure 5.6: Minimizers and their cross sections for genus = 2 and $v_0 = 0.66(cont'd)$

The Canham problem seems to have unique solution outside the above interval. There are two types of minimizers (D_{3h} and D_{2h} surface in [20]).

5.2 Helfrich problem

The solution of the Helfrich problem (with three constraints) exhibits a variety of interesting shapes when the constraint values v_0, m_0 vary. See Figure 5.7.

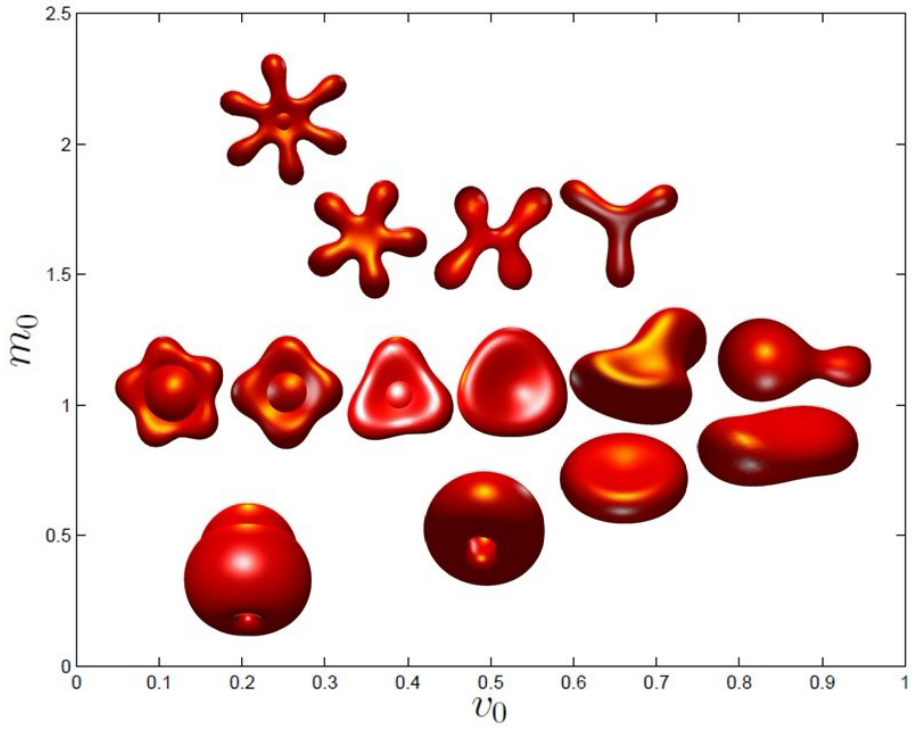
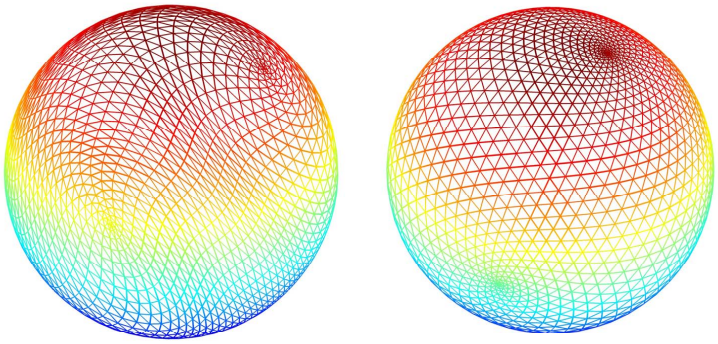


Figure 5.7: Solutions to Helfrich problem for genus = 0

5.3 Harmonic energy

For subdivision surfaces, different control meshes may yield the same shape of surfaces. For example, Figure 5.8 are two round spheres, as their Willmore energies are both close to $4\pi \approx 12.56637$ and the only genus 0 surface with $W = 4\pi$ is a sphere. However, their control meshes are different: the right mesh consists of equilateral triangles, while the left mesh does not.

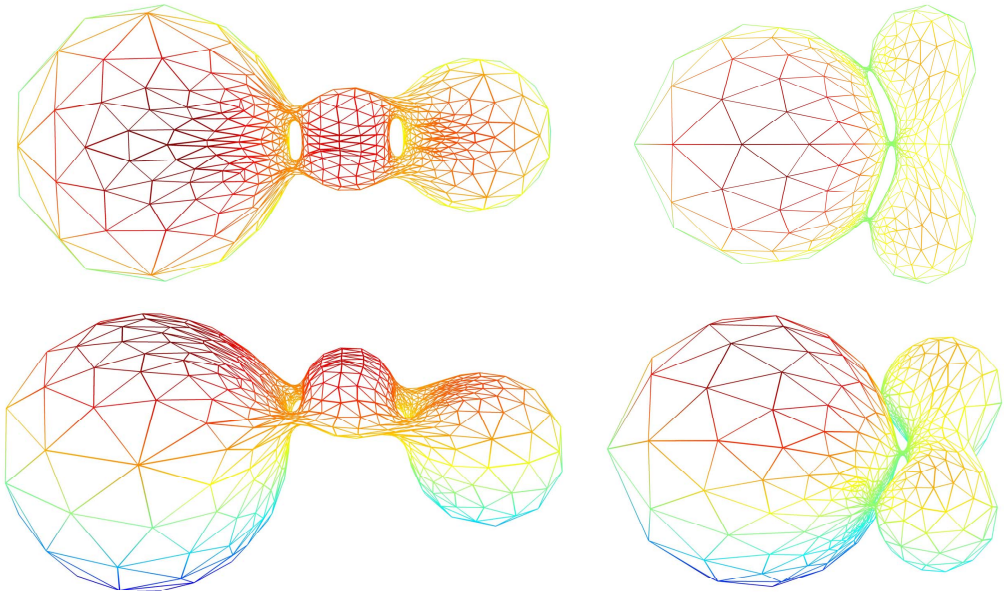


(a) $W = 12.56637$ (b) $W = 12.56641$

Figure 5.8: Spheres with non-equilateral and equilateral triangular meshes: (a) obtained by minimizing Willmore energy, (b) obtained by harmonic map

We consider the right mesh "better" since it is more uniform and provides a better parametrization. The way we quantify how good a parametrization is via *Harmonic Energy*. Figure 5.8(b) is the obtained by the harmonic map, which maps a surface to a surface with the lowest harmonic energy.

We will not explain here why harmonic energy quantifies if a mesh provides a good parametrization or not, but show an example to illustrate how harmonic energy can help regularize the parametrization and find the true minimizer in the Helfrich problem. By adding the Harmanic Energy to the objective function as a regularization term, it helps to find a surface with a lower Willmore energy in Figure 5.9.



(a) $W = 34.31321$ (b) $W = 33.00928$

Figure 5.9: Solutions to the Helfrich problem for genus=2 with $v_0 = 0.7, m_0 = 1.5\times 4\pi$. (a) Without HE as regularization term (b) With HE as regularization term

5.4 Other numerical results

(I) Solutions to the Canham problem for $v_0 = 0.1$

We conjecture that the solution to the Canham problem for $v_0 = 0$ for any genus is a double sphere with $W = 8\pi \approx 25.13274$. We show the minimizers and their Willmore energy obtained from our solver for genus = 1, 2 and 3. See Figure 5.11-Figure 5.14.

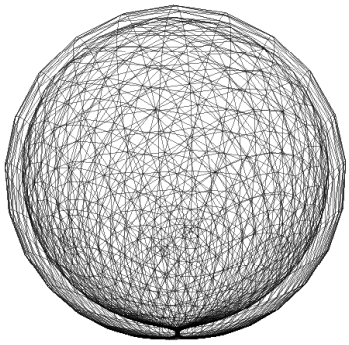


Figure 5.10: genus0, $v_0 = 0.1, W = 25.13161$

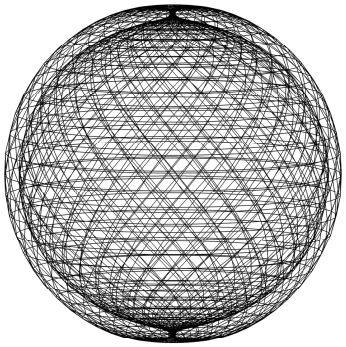


Figure 5.11: genus1, $v_0 = 0.1, W = 25.12887$

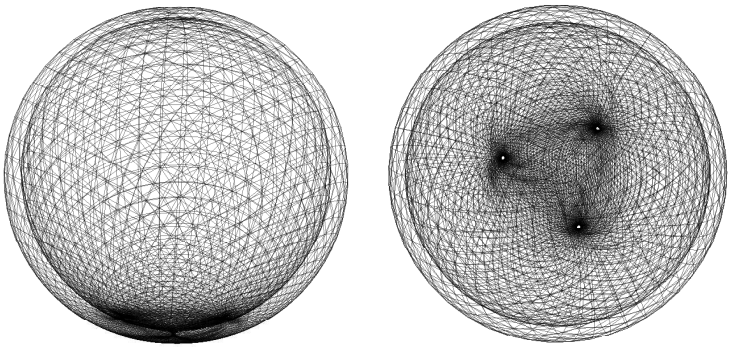


Figure 5.12: genus2, $v_0 = 0.1, W = 25.13174$

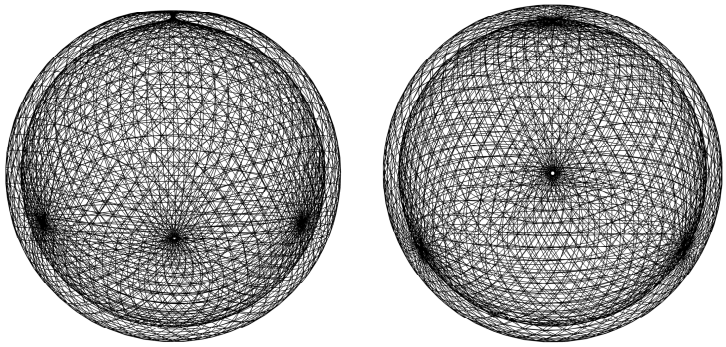


Figure 5.13: genus3-S4, $v_0 = 0.1, W = 25.14411$

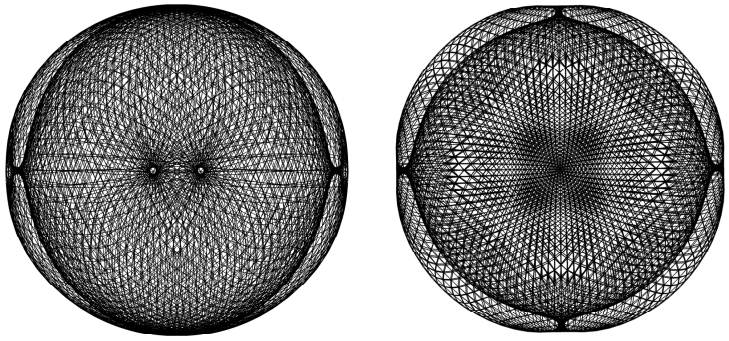


Figure 5.14: genus3-D4, $v_0 = 0.1, W = 25.27570$

Notice for genus $= 3$, the double sphere with S4 symmetry handles has a significant lower Willmore energy than the one with D4 symmetry handles. This seems to imply that S4 is the correct symmetry for the true minimizer.

(II) Solutions to the Willmore problem for genus $= 3$

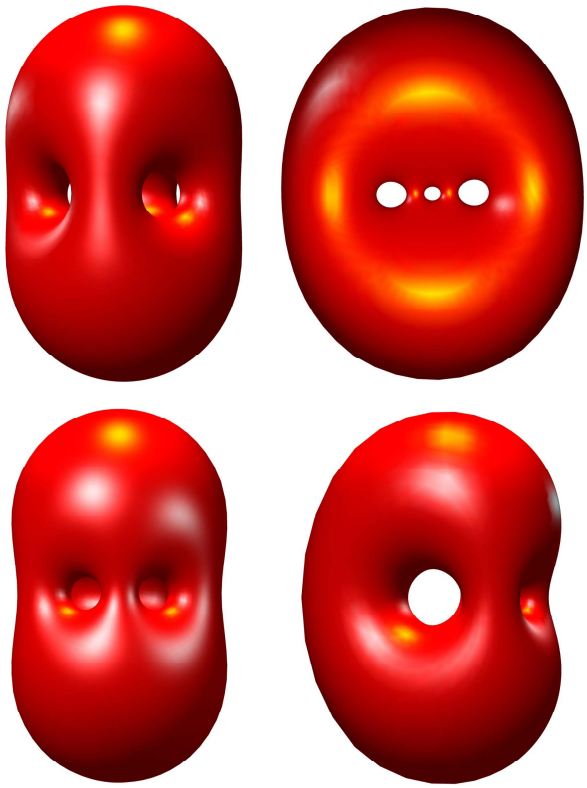


Figure 5.15: Minimizers to Willmore problem for genus = 3

6. Uniqueness of solution for the Canham problem, $v_0 \in [v_{\mathrm{SC}}, 1)$

Computational results suggest that in many situations the solution of either the Canham or Helfrich problem is unique up to homothetic transformations in $\mathbb{R}^3$. As discussed in the introduction chapter, this is puzzling in virtue of the following hasty dimensionality count: Recall that the Willmore energy is invariant under sphere inversions; and there are three degrees of freedom in choosing the center of a sphere. Given any minimizer of a Canham or Helfrich problem, one can apply to it the three dimensional family of sphere inversions and expect to have enough degrees of freedom to satisfy the reduced volume constraint ($v = v_0$) or reduced volume plus mean curvature constraints ($v = v_0$, $m = m_0$), yielding a two- or one-parameter family of solutions.

This dimensional count would sound more convincing when the genus is 2 or above, thanks to the Hurwitz's automorphisms theorem. But even so, there appears to be regions in the v_0-m_0 plane where uniqueness prevails; see, for example, the claim in [20]. We are not aware of any mathematically rigorous uniqueness result.

A natural starting point to address the uniqueness problem is to study the Canham problem with the reduced volume constraint in the interval

$$v_0 \in [v_{\mathrm{SC}}, 1),$$

in which we know from the Marques-Neves theorem that the solution is a Clifford torus.

Let $T := T_R := \{(R + \cos u)\cos v, (R + \cos u)\sin v, \sin u) : u, v \in [0, 2\pi]\}$, a torus with major radius $R \in (1, \infty)$, minor radius 1, and the z-axis as the axis of revolution. Let $i_{(x,y,z)}$ be the inversion map about the unit sphere centered at (x, y, z) of $\mathbb{R}^3$. Our

goal is to characterize all the Euclidean shapes of the (torodial) cyclides, i.e. we would like to find a parameterization of the 'shape space'

$$\left\{ i_{(x,y,z)}(T_R) : R > 1,\ (x,y,z) \in \mathbb{R}^3 \backslash T_R \right\} \Big/ \mathrm{Hom}(3), \tag{6.1}$$

and also one specifically for the Clifford tori:

$$\left\{ i_{(x,y,z)}(T_{\sqrt{2}}) : (x,y,z) \in \mathbb{R}^3 \backslash T_{\sqrt{2}} \right\} \Big/ \mathrm{Hom}(3). \tag{6.2}$$

Here '/Hom(3)' means we identify two cyclides if they can be transformed from one to another by a homothety in $\mathbb{R}^3$. Since we are primarily interested in Euclidean shapes here, we avoid sphere inversions centered at points on T_R itself.

For a reason that will become clear soon, we work with cylindrical coordinates $(\rho\cos(\theta), \rho\sin(\theta), z) = (x, y, z)$. For any $\varrho > 0$, let $\mathcal{C}(\varrho) = \mathcal{C}(\varrho; R)$ be the circle in the ϱ-z plane with a diameter connecting $(\varrho, 0)$ and $((R^2-1)/\varrho, 0)$; see Figure 6.4. By convention, $\mathcal{C}(0) = \mathcal{C}(\infty)$ is the z-axis. In general, we have $C(\varrho) = C((R^2-1)/\varrho)$.

These circles on the plane can be extended to the following tori in 3-D:

$$\mathcal{T}(\varrho) := \mathcal{T}(\varrho; R) := \{(\rho\cos(\theta), \rho\sin(\theta), z) : (\rho, z) \in \mathcal{C}(\varrho),\ \theta \in [0, 2\pi]\}. \tag{6.3}$$

For any fixed R, the torus $\mathcal{T}(\varrho)$ lies completely outside, on, or inside the torus T when $\varrho \in [0, R-1) \cup (R+1, \infty]$, $\varrho = R \mp 1$, or $\varrho \in (R-1, R+1)$, respectively. In particular, $\mathcal{T}(R \pm 1; R) = T_R$. On the ρ-z plane, these correspond to the red, green and blue circles in Figure 6.4. While the one-parameter family of circles $\{\mathcal{C}(\varrho) : \varrho \in [0, \sqrt{R^2-1}]\}$ partitions the ρ-z plane,[1] the corresponding one-parameter family of tori $\{\mathcal{T}(\varrho)\}$

[1] For any (ρ, z), it lies on the circle $\mathcal{C}(\varrho^+) = \mathcal{C}(\varrho^-)$, where

$$\varrho^{\pm} = \frac{\left(\rho^2 + z^2 + R^2 - 1\right) \pm \sqrt{\left(\rho^2 + z^2 + R^2 - 1\right)^2 - 4\rho^2(R^2-1)}}{2\rho}.$$

partitions $\mathbb{R}^3$.

Maxwell's characterization of a cyclide. It is well-known that any (torodial) cyclide $\mathfrak{C}$ has two orthogonal planes of mirror symmetry (see, for example, [25, 6, 10].) We make the observation that the three-dimensional shape of a toroidal cyclide $\mathfrak{C}$ is uniquely determined by certain measurements of the cross section of $\mathfrak{C}$ on either one of the two symmetry planes.

We use Maxwell's characterization of cyclides [25, 6, 10]: any cyclide $\mathfrak{C}$ is the envelope of all the spheres centered at the points P on a given ellipse $\mathcal{E}$ with radii $r(P)$, $P \in \mathcal{E}$, satisfying $r(P) + \overline{FP} = L$, where F is one of the foci of $\mathcal{E}$ and L is a constant in a suitable range. We can think of L as the length of a taut string attached in one end to F; the string slides smoothly on $\mathcal{E}$ and traces out spheres with the other end. See Figure 6.1. This characterization suggests that the shape of $\mathfrak{C}$ can be characterized by the ratio

$$a : f : L,$$

where a, f and L are the major radius of $\mathcal{E}$, the focal length of $\mathcal{E}$, and the length of the string, respectively.[2]

The major axis of $\mathcal{E}$ lies on the intersecting line of the two symmetry planes of $\mathfrak{C}$. In the following, P_1 refers to the symmetry plane where $\mathcal{E}$ lies, whereas P_2 ($\perp P_1$) refers to the other symmetry plane. The cross section $\mathfrak{C} \cap P_1$ consists of two circles **exterior to each other**, whereas the cross section $\mathfrak{C} \cap P_2$ consists of two circles with **one lying inside the other** (see Figure 6.3).

Denote the radii of the two circles in $\mathfrak{C} \cap P_1$ by r_1 and r_2 and the distance between the two centers by d (see Figure 6.3). Similarly, let $\tilde{r}_1$ and $\tilde{r}_2$ be the radii of the two circles in $\mathfrak{C} \cap P_2$ and $\tilde{d}$ be the distance between the two centers. By convention,

[2]This already explains why the shape space (6.1) is two-dimensional.

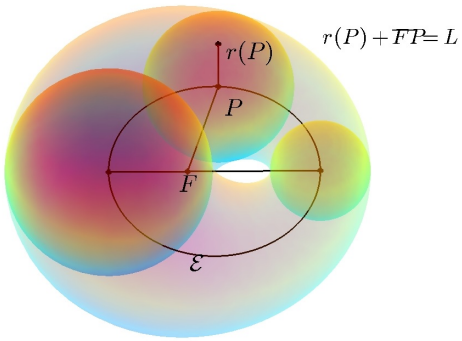


Figure 6.1: Maxwell's characterization of a ring cyclide

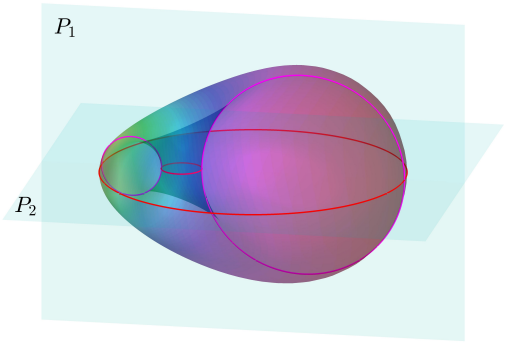
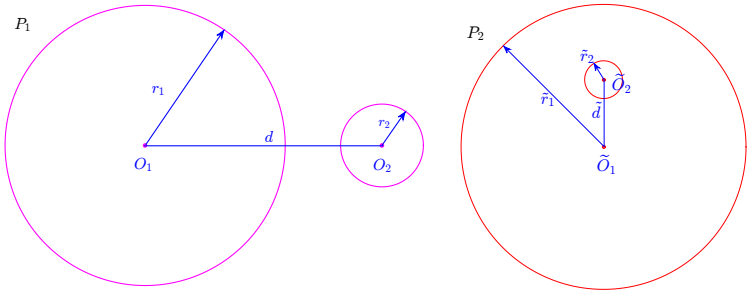


Figure 6.2: Two planes of mirror symmetry



(a) $i_{(x,y,z)}(T_R) \cap P_1$ (b) $i_{(x,y,z)}(T_R) \cap P_2$

Figure 6.3: Cross sections of $i_{(x,y,z)}(T_R)$ at P_1 and P_2

$r_1 \geqslant r_2$, $\tilde{r}_1 \geqslant \tilde{r}_2$. The three sets of measurements (r_1, r_2, d), $(\tilde{r}_1, \tilde{r}_2, \tilde{d})$ and (a, f, L) of a cyclide $\mathfrak{C}$ are related by the following equations:

$$a = \frac{d}{2}, \quad f = \frac{r_1 - r_2}{2}, \quad L = \frac{d + r_1 + r_2}{2}. \tag{6.4}$$

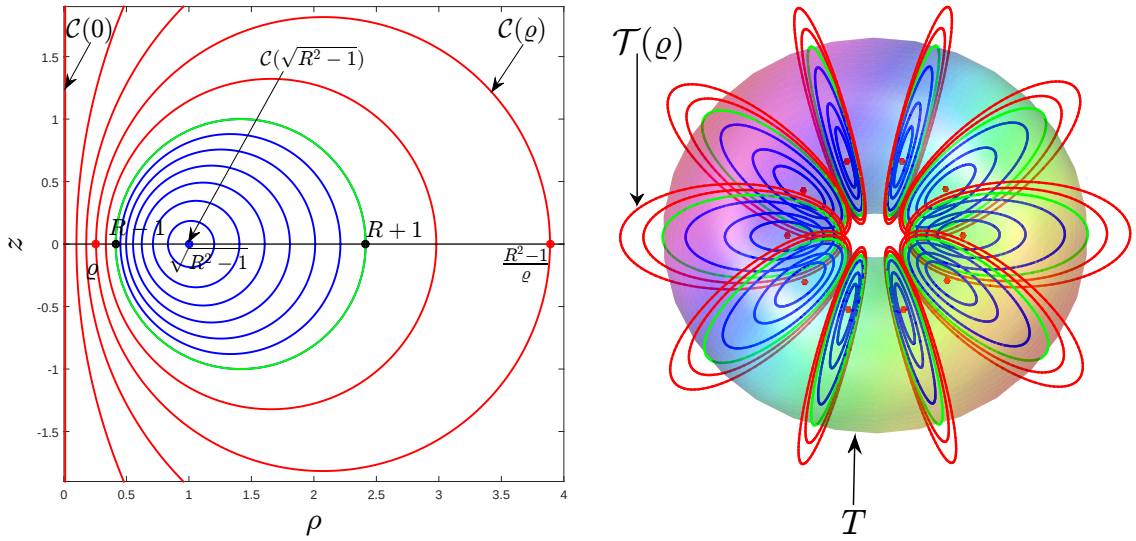
$$\tilde{r}_1 = \frac{d + (r_1 + r_2)}{2}, \quad \tilde{r}_2 = \frac{d - (r_1 + r_2)}{2}, \quad \tilde{d} = r_1 - r_2. \tag{6.5}$$

Since the maps $(a, f, L) \mapsto (r_1, r_2, d)$ and $(r_1, r_2, d) \mapsto (\tilde{r}_1, \tilde{r}_2, \tilde{d})$ are both linear and invertible, and that $a{:}f{:}L$ determines the three-dimensional shape of $\mathfrak{C}$, we conclude

that

$$r_1 : r_2 : d \quad \text{and} \quad \tilde{r}_1 : \tilde{r}_2 : \tilde{d} \tag{6.6}$$

can also be used to determine the three-dimensional shape of the cyclide $\mathfrak{C}$.



(a) (b)

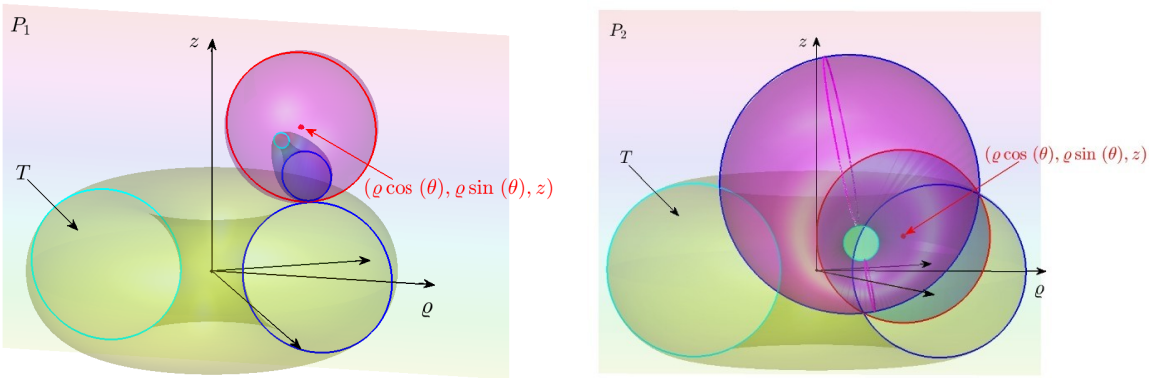
Figure 6.4: (a) $\mathcal{C}(\varrho)$ for various values of $\varrho \in (0, R-1)$ (in red) and $\varrho \in (R-1, \sqrt{R^2-1})$ (in blue). Note that $\mathcal{C}(\varrho) = \mathcal{C}((R^2-1)/\varrho)$ and $\mathcal{C}(\sqrt{R^2-1})$ degenerates into a point. (b) $\mathcal{T}(\varrho)$: for visualization purpose we only display the part of $\mathcal{T}(\varrho)$ corresponding to ten equally-spaced θ's in (6.3) in order not to occlude the underlying torus T.

Theorem 6.0.1. *For any fixed $R \in (1, \infty)$ and $\varrho \in [0, \infty] \backslash \{R-1, R+1\}$, all the cyclides in*

$$\Big\{ i_{(x,y,z)}(T_R) : \quad (x, y, z) \in \mathcal{T}(\varrho; R) \Big\}, \tag{6.7}$$

are homothetic in $\mathbb{R}^3$.

Proof: Let P be the plane determined by the z-axis and (x, y, z). It can be shown



(a) (b)

Figure 6.5: $i_{(\varrho\cos(\theta),\varrho\sin(\theta),z)}(T_R)$: (a) $(\varrho\cos(\theta), \varrho\sin(\theta), z)$ is outside T_R, (b) $(\varrho\cos(\theta), \varrho\sin(\theta), z)$ is inside T_R

that

$$P = \begin{cases} \text{the } P_1 \text{ of } i_{(x,y,z)}(T_R), & \text{when } (x, y, z) \text{ lies outside of } T_R \\ \text{the } P_2 \text{ of } i_{(x,y,z)}(T_R), & \text{when } (x, y, z) \text{ lies inside of } T_R. \end{cases} \tag{6.8}$$

See Figure 6.5. Since we already explained that the three dimensional shape of the cyclide $i_{(x,y,z)}T_R$ is determined by the measurements of the two dimensional cross section as in (6.6), we can prove the theorem by showing that all the circle pairs in

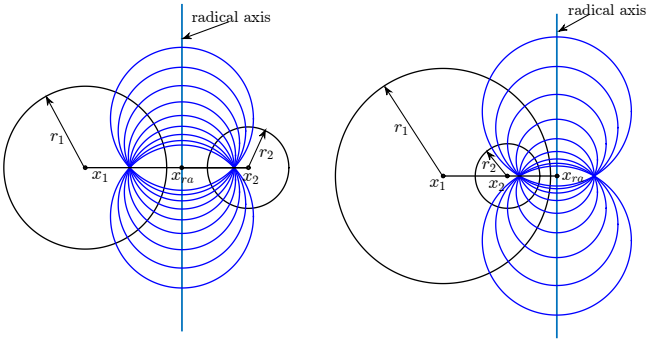
$$\Big\{ i_{(\rho,z)}(T_R \cap P) : \quad (\rho, z) \in \mathcal{C}(\varrho) \Big\}, \tag{6.9}$$

are homothetic in $\mathbb{R}^2$.[3] (Here $i_{(\rho,z)}$ stands for a circle inversion with respect to the

[3]Here and below, we think of any such circle pair as a 2-dimensional shape by itself. Two such circle pairs are homothetic if they share the same ratio $r_1{:}r_2{:}d$ (in the case of Figure 6.3(a)) or $\tilde{r}_1{:}\tilde{r}_2{:}\tilde{d}$ (in the case of Figure 6.3(b)).

unit circle centered at (ρ, z).)

We prove this 2-dimensional result in two steps.



(a) (b)

Figure 6.6: Radical Axis of a circle pair: (a) one circle inside the other; (b) one circle contains the other. The blue circles meet the circle pair orthogonally.

(I) In general, if we invert two circles centered at $(x_1, 0)$ and $(x_2, 0)$ with radii r_1 and r_2 about a circle centered anywhere on the line $\left\{(x_{ra}, y) | x_{ra} = \dfrac{(x_2^2 - x_1^2) + (r_1^2 - r_2^2)}{2(x_2 - x_1)}\right\}$, the resulting circle pair is homothetic to the original circle pair. This line is called the *radical axis* of the original circle pair; see Figure 6.6.

Recall P is the ρ-z plane. Now let us consider the image of the cross section $T \cap P$ under the inversion $i_{(\varrho,0)}$. It is not hard to check that the resulting circle pair has the following centers and radii:

$$\begin{aligned} \bar{r}_1 &= \frac{1}{|(\varrho - R)^2 - 1|}, \ \bar{r}_2 = \frac{1}{|(\varrho + R)^2 - 1|} \\ \bar{O}_1 &= \left(\varrho - \frac{\varrho - R}{(\varrho - R)^2 - 1}, 0\right), \ \bar{O}_2 = \left(\varrho - \frac{\varrho + R}{(\varrho + R)^2 - 1}, 0\right) \end{aligned} \tag{6.10}$$

As O_1 and O_2 lie on the ρ-axis, the radical axis of the circle pair in the cross

section $i_{(\varrho,0)}(T_R \cap P)$ is parallel to the z-axis, with ρ-coordinate

$$\rho_{ra} = \varrho - \frac{\varrho}{\varrho^2 + 1 - R^2}.$$

Now the circle pairs in

$$\left\{ i_{(\rho_{ra},z)} \circ i_{(\varrho,0)}(T_R \cap P) : \ z \in \mathbb{R} \right\}$$

are all homothetic. The theorem is proved if we prove that every circle pair in $\left\{ i_{(\varrho,z)}(T_R \cap P) : \ (\varrho, z) \in \mathcal{C}(\varrho) \right\}$ is homothetic to some circle pair in $\left\{ i_{(\rho_{ra},z)} \circ i_{(\varrho,0)}(T_R \cap P) : \ z \in \mathbb{R} \right\}$. We do so in the second step of the proof.

(II) Since an arbitrary composition of conformal transformations can be written as a composition of an inversion (of radius 1) with a homothety (See page 92 of [2]),

$$i_{(x_{ra},z)} \circ i_{(\varrho,0)} = \mathcal{H} \circ i_{(\varrho_1,z_1)}. \tag{6.11}$$

We can use the following properties of an inversion i_O to find (ϱ_1, z_1):

- $i_O(O) = \infty$;
- $i_O(\infty) = O$;
- $i_O(Q_1) = Q_2$ is equivalent to $i_O(Q_2) = Q_1$.

By the first property, we have

$$i_{(\rho_{ra},z)} \circ i_{(\varrho,0)}((\varrho_1, z_1)) = \mathcal{H} \circ i_{(\varrho_1,z_1)}((\varrho_1, z_1)) = \infty.$$

Apply the second property, we obtain

$$i_{(\varrho,0)}((\varrho_1, z_1)) = (\rho_{ra}, z).$$

By the third property,

$$(\varrho_1, z_1) = i_{(\varrho,0)}(\rho_{ra}, z),$$

This means the set of all (ϱ_1, z_1) in (6.11) is the image of the line $\{(\rho_{ra}, z)|z \in \mathbb{R}\}$ under the inversion $i_{(\varrho,0)}$, which is a circle. It is not hard to see that this circle has a diameter on the ρ-axis. One end of the diameter is $i_{(\varrho,0)}((\rho_{ra}, \infty)) = (\varrho, 0)$, and the other end is $i_{(\varrho,0)}((\varrho_{ra}, 0)) = \frac{R^2-1}{\varrho}$. The circle is $\mathcal{C}(\varrho)$. ■

In virtue of this result, we introduce the notation $i_\varrho(T_R)$ to represent the common Euclidean shape of the cyclides in (6.7). Formally, $i_\varrho(T_R)$ is an element in the shape space (6.1). In the following lemma, we are going to show that different values of R may yield the same shape of cyclides under different inversions.

Lemma 6.0.2. *Let $r_1 : r_2 = \lambda$, $\tilde{r}_1 : \tilde{r}_2 = \tilde{\lambda}$. For any $R \in (1, \infty)$, the cross sections $i_\varrho(T_R) \cap P_1$ and $i_\varrho(T_R) \cap P_2$ have the following measurements:*

(i) when $\varrho \in [0, R-1)$ (corresponding to the red circles in Figure 6.4(a)),

$$\begin{aligned} r_1 : r_2 : d &= \lambda : 1 : \sqrt{(\lambda-1)^2 + 4\lambda R^2}, \\ \tilde{r}_1 : \tilde{r}_2 : \tilde{d} &= \tilde{\lambda} : 1 : \sqrt{(\tilde{\lambda}+1)^2 - 4\tilde{\lambda}\frac{R^2}{R^2-1}}; \end{aligned} \tag{6.12}$$

(ii) when $\varrho \in (R-1, \sqrt{R^2-1}]$ (corresponding to the blue circles in Figure 6.4(a)),

$$\begin{aligned} r_1 : r_2 : d &= \lambda : 1 : \sqrt{(\lambda-1)^2 + 4\lambda\frac{R^2}{R^2-1}}, \\ \tilde{r}_1 : \tilde{r}_2 : \tilde{d} &= \tilde{\lambda} : 1 : \sqrt{(\tilde{\lambda}+1)^2 - 4\tilde{\lambda}R^2}, \end{aligned} \tag{6.13}$$

Proof: (i) When $\varrho \in [0, \sqrt{R^2-1})$, the cross section $i_{(\varrho,0)}(T_R \cap P)$ (Figure 6.3(a)) is given by:

$$O_1 = \left(\varrho - \frac{(\varrho - R)}{(\varrho - R)^2 - 1}, 0\right), \quad O_2 = \left(\varrho - \frac{(\varrho + R)}{(\varrho + R)^2 - 1}, 0\right), \tag{6.14}$$

$$r_1 = \frac{1}{(\varrho - R)^2 - 1}, \quad r_2 = \frac{1}{(\varrho + R)^2 - 1}, \tag{6.15}$$

$$d = \frac{2R(R^2 - 1 - \varrho^2)}{\big((\varrho - R)^2 - 1\big)\big((\varrho + R)^2 - 1\big)}. \tag{6.16}$$

Since our goal is to express the ratio $r_1 : r_2 : d$ in terms of r_1 and r_2, we first rewrite ϱ in terms of r_1 and r_2. By (6.15),

$$\frac{(\varrho + R)^2 - 1}{(\varrho - R)^2 - 1} = \frac{r_1}{r_2}.$$

Thus ϱ can be solved as follows:

$$\varrho^2 - 2\frac{r_1 + r_2}{r_1 - r_2} R\varrho + (R^2 - 1) = 0$$

$$\varrho = \frac{r_1 + r_2}{r_1 - r_2} R \pm \sqrt{\frac{4r_1 r_2 R^2}{(r_1 - r_2)^2} + 1}$$

Substitute ϱ, we get

$$\begin{aligned} d &= |O_1 - O_2| \\ &= \Big|\frac{(\rho - R)}{(\rho - R)^2 - 1} - \frac{(\rho + R)}{(\rho + R)^2 - 1}\Big| \\ &= |r_1(\rho - R) - r_2(\rho + R)| \\ &= |(r_1 - r_2)\rho - (r_1 + r_2)R| \\ &= \sqrt{4r_1 r_2 R^2 + (r_1 - r_2)^2}. \end{aligned} \tag{6.17}$$

Next we analyze the ratio $\tilde{r}_1 : \tilde{r}_2 : \tilde{d}$. By (6.5),

$$\tilde{r}_1 + \tilde{r}_2 = d. \tag{6.18}$$

Together with (6.17), we have

$$(\tilde{r}_1 + \tilde{r}_2)^2 = 4r_1r_2R^2 + (r_1 - r_2)^2. \tag{6.19}$$

Again by (6.5),

$$(\tilde{r}_1 - \tilde{r}_2)^2 = (r_1 + r_2)^2. \tag{6.20}$$

Subtract (6.20) from (6.19) and solve for r_1r_2, we obtain

$$r_1r_2 = \frac{\tilde{r}_1\tilde{r}_2}{R^2 - 1}. \tag{6.21}$$

Combine (6.5), (6.20) and (6.21), we can finally express $\tilde{d}$ in terms of $\tilde{r}_1$ and $\tilde{r}_2$

$$\begin{aligned} \tilde{d}^2 &= (r_1 - r_2)^2 \\ &= (r_1 + r_2)^2 - 4r_1r_2 \\ &= (\tilde{r}_1 - \tilde{r}_2)^2 - 4\frac{\tilde{r}_1\tilde{r}_2}{R^2 - 1} \\ &= (\tilde{r}_1 + \tilde{r}_2)^2 - 4\tilde{r}_1\tilde{r}_2\frac{R^2}{R^2 - 1}. \end{aligned} \tag{6.22}$$

(ii) When $\varrho \in (R - 1, \sqrt{R^2 - 1}]$, the cross section (Figure 6.3(b))are given by:

$$\widetilde{O}_1 = \left(\rho - \frac{(\rho - R)}{(\rho - R)^2 - 1}, 0\right), \quad \widetilde{O}_2 = \left(\rho - \frac{(\rho + R)}{(\rho + R)^2 - 1}, 0\right) \tag{6.23}$$

$$\tilde{r}_1 = \frac{1}{1 - (\rho - R)^2}, \quad \tilde{r}_2 = \frac{1}{(\rho + R)^2 - 1}, \tag{6.24}$$

$$\tilde{d} = \frac{2R(R^2 - 1 - \rho^2)}{\big(1 - (\rho - R)^2\big)\big((\rho + R)^2 - 1\big)}. \tag{6.25}$$

(ii) can be proved similarly as (i) by (6.23)-(6.25) and the following equations:

$$r_1 = \frac{\tilde{r}_1 - \tilde{r}_2 + \tilde{d}}{2}, \quad r_2 = \frac{\tilde{r}_1 - \tilde{r}_2 - \tilde{d}}{2}, \quad d = \tilde{r}_1 + \tilde{r}_2. \tag{6.26}$$

■

Lemma 6.0.3. *For any $R \in (1, \infty)$, $\varrho \in [0, \sqrt{R^2-1}]$,*

$$i_\varrho(T_R) = i_{\varrho'}(T_{R'})$$

where

$$(R', \varrho') = \frac{1}{\sqrt{R^2-1}}\Big(R, \frac{\sqrt{R^2-1}-\varrho}{\sqrt{R^2-1}+\varrho}\Big). \tag{6.27}$$

Proof: (I) Denote $R' = \dfrac{R}{\sqrt{R^2-1}}$. We first prove that for any $\varrho \in [0, \sqrt{R^2-1}]$, there exists a $\varrho' \in [0, \sqrt{R'^2-1}]$, s.t.

$$i_\varrho(T_R) = i_{\varrho'}(T_{R'})$$

We will show how to choose such a ϱ' in part (II).

We will prove in the next theorem that $r_1 : r_2$ increases as $\varrho \in [0, R-1)$ and decreases as $\varrho \in (R-1, \sqrt{R^2-1}]$. Moreover, the ratio $r_1 : r_2$ spans$[1, \infty)$ when ϱ varies in both intervals. Thus for any fixed $\lambda \in [1, \infty)$, there exists $\varrho \in [0, R-1)$, s.t. the radii of the cross section $i_\varrho(T_R)$ satisfy $r_1 : r_2 = \lambda$; there also exists $\varrho \in (R-1, \sqrt{R^2-1}]$, s.t. the above condition also holds. These two maps are one-to-one because of the monotonicity we just mentioned. Therefore, in the following proof, instead of fixing $\varrho \in [0, \sqrt{R^2-1}]$, we just fix λ.

By the fact we mentioned above and Lemma 6.0.2 (i), for any given λ, there exists a $\varrho \in [0, R-1)$, s.t. $i_{(\varrho,0)}(T_R \cap P_1)$ has the following measurements

$$\begin{aligned} r_1 : r_2 &= \lambda, \\ r_1 : r_2 : d &= \lambda : 1 : \sqrt{(\lambda-1)^2 + 4\lambda R^2}. \end{aligned} \tag{6.28}$$

For the same reason, there exists a $\varrho' \in (R'-1, \sqrt{R'^2-1}]$, s.t. $i_{(\varrho',0)}(T_{R'} \cap P_1)$ has the following measurements

$$\begin{aligned} r_1 : r_2 &= \lambda, \\ r_1 : r_2 : d &= \lambda : 1 : \sqrt{(\lambda-1)^2 + 4\lambda \frac{R'^2}{R'^2-1}} \\ &= \lambda : 1 : \sqrt{(\lambda-1)^2 + 4\lambda R^2}. \end{aligned} \tag{6.29}$$

The last equal sign is due to the fact $\dfrac{R'^2}{R'^2-1} = R^2$. Notice (6.28) and (6.29) are the same. Thus

$$i_\varrho(T_R) = i_{\varrho'}(T_{R'}).$$

Similarly, for the given λ, there exist another $\varrho \in (R-1, \sqrt{R^2-1}]$ and $\varrho' \in [0, R'-1)$, s.t.

$$i_\varrho(T_R) = i_{\varrho'}(T_{R'}).$$

Combine the above two results, the first part of lemma is proved: For any $R \in (1, \infty)$, $\varrho \in [0, \sqrt{R^2-1}]$,

$$i_\varrho(T_R) = i_{\varrho'}(T_{R'}).$$

(II) In this part, we are going to prove when

$$\varrho' = \frac{1}{\sqrt{R^2-1}} \frac{\sqrt{R^2-1} - \varrho}{\sqrt{R^2-1} + \varrho},$$

$i_\varrho(T_R)$ and $i_{\varrho'}(T_{R'})$ have the same $r_1 : r_2$. We prove it is true for $\varrho \in [0, R-1)$, and the proof is similar for $\varrho \in (R-1, \sqrt{R^2-1}]$.

In the previous proof, we were more interested in the shape of a cyclide itself. Therefore, we chose r_1 and r_2 as our parameters to describe the shape of a cyclide (for example Lemma 6.0.2), as they are more geometric and intuitive. However, now

we need to analyze shapes of all cyclide $i_\varrho(T_R)$ for different ϱ's and R's. From now on, we consider ϱ and R as our parameters and denote $r_1(R,\varrho)$ and $r_2(R,\varrho)$ as the radii of the cross section $i_{(\varrho,0)}(T_R \cap P_1)$.

In (I) we choose ϱ and ϱ', s.t. $i_\varrho(T_R)$ and $i_{\varrho'}(T_{R'})$ have the same ratio

$$\frac{r_1(R,\varrho)}{r_2(R,\varrho)} = \frac{r_1(R',\varrho')}{r_2(R',\varrho')} = \lambda.$$

Next we are going to calculate the above two ratios to solve for ϱ'. By (6.15),

$$\frac{r_1(R,\varrho)}{r_2(R,\varrho)} = \frac{(\varrho+R)^2-1}{(\varrho-R)^2-1}, \text{ where } \varrho \in [0, R-1). \tag{6.30}$$

Combine (6.24) - (6.26), we obtain the following two radii of the cross section $i_\varrho(T_R)$

$$r_1(R,\varrho) = \frac{R-1}{\varrho^2-(R-1)^2}, \quad r_2(R,\varrho) = \frac{R+1}{(R+1)^2-\varrho^2}, \tag{6.31}$$

where $\varrho \in (R-1, \sqrt{R^2-1}]$.

Thus

$$\frac{r_1(R',\varrho')}{r_2(R',\varrho')} = \frac{(R'-1)\Big((R'+1)^2-\varrho'^2\Big)}{(R'+1)\Big(\varrho'^2-(R'-1)^2\Big)},$$

for $\varrho' \in (R'-1, \sqrt{R'^2-1}]$.

Solve the equation

$$\frac{r_1(R,\varrho)}{r_2(R,\varrho)} = \frac{r_1(R',\varrho')}{r_2(R',\varrho')},$$

we obtain

$$\varrho' = \frac{1}{\sqrt{R^2-1}} \frac{\sqrt{R^2-1}-\varrho}{\sqrt{R^2-1}+\varrho}.$$

■

Theorem 6.0.4. *The shape space (6.1) is in one-to-one correspondence with the set of $i_\varrho(T_R)$, where (R,ϱ) satisfies $R \in (1,\sqrt{2})$, $\varrho \in [0,R-1) \cup (R-1, \sqrt{R^2-1}]$ or*

$R = \sqrt{2},\ \varrho \in [0, \sqrt{2} - 1)$.

Proof: (I) We first prove for any fixed R, $\dfrac{r_1(R,\varrho)}{r_2(R,\varrho)}$ increases as $\varrho \in [0, R-1)$ and decreases as $\varrho \in (R-1, \sqrt{R^2-1}]$, and the vaule of the ratio spans$[1, \infty)$ when ϱ varies in both intervals (the fact we stated in Lemma 6.0.3 (I)).

Take the derivative of (6.30) w.r.t. ϱ, we obtain

$$\Big(\frac{r_1(R,\varrho)}{r_2(R,\varrho)}\Big)' = \frac{4R(R^2-\varrho^2-1)}{\big((\varrho-R)^2-1\big)^2} > 0,$$

as $\varrho \in [0, R-1)$. Moreover,

$$\frac{r_1(R,0)}{r_2(R,0)} = 1 \text{ and } \frac{r_1(R,R-1)}{r_2(R,R-1)} = \infty.$$

Thus, the fact is true for $\varrho \in [0, R-1)$ and we can prove it similarly for $\varrho \in (R-1, \sqrt{R^2-1}]$.

(II) For any fixed R, the ratio of the measurements of $i_\varrho(T_R)$

$$r_1 : r_2 : d = \lambda : 1 : \sqrt{(\lambda-1)^2 + 4\lambda R^2},$$

is different when ϱ varies in $[0, R-1)$, since $r_1{:}r_2 = \lambda$ is different.

Similarly, the ratio of the measurements of $i_\varrho(T_R)$

$$r_1 : r_2 : d = \lambda : 1 : \sqrt{(\lambda-1)^2 + 4\lambda \frac{R^2}{R^2-1}},$$

is different when ϱ varies in $(R-1, \sqrt{R^2-1}]$.

Therefore, the cyclides in $\{i_\varrho(T_R), \varrho \in [0, R-1)\}$ and $\{i_\varrho(T_R), \varrho \in (R-1, \sqrt{R^2-1}]\}$ are all uniquely determined by ϱ. Moreover, these two sets are disjoint when $R^2 \neq \dfrac{R^2}{R^2-1}$, i.e. $R \neq \sqrt{2}$ and they are identical when $R = \sqrt{2}$.

Now we analyze the shape space (6.1) for different R's. Let $R \neq R'$, $i_\varrho, (T_R)$ and $i_{\varrho'}(T_{R'})$ have the same shape if and only if

i) Their λ's are the same.

This can always be achieved due to (I) and more specifically we just need to choose ϱ' to be the ϱ' in (6.27).

ii) $R = \dfrac{R}{R^2 - 1}$.

$R \mapsto \dfrac{R}{R^2 - 1}$ has a fixed point at $R = \sqrt{2}$ and it maps $[0, \sqrt{2})$ to $(\sqrt{2}, \infty)$. This map is one to one on $[0, \sqrt{2})$.

Therefore, the shape space of

$$\{i_\varrho(T_R) : R \in (1, \sqrt{2}), \varrho \in [0, R-1) \cup (R-1, \sqrt{R^2-1}]\}$$

is the same with that of

$$\{i_\varrho(T_R) : R \in (\sqrt{2}, \infty), \varrho \in [0, R-1) \cup (R-1, \sqrt{R^2-1}]\}.$$

■

Conjecture 6.0.5. *i) The reduced volume v_0 of $i_\varrho(T_R)$ is increasing w.r.t. $\varrho \in [0, R-1)$.*

ii) The reduced volume v_0 of a cyclide is increasing w.r.t. $\dfrac{r_1}{r_2}$, where r_1 and r_2 are as shown in Figure 6.3 (a).

ii) follows immediately if we combine i) and the fact that $\dfrac{r_1}{r_2}$ is increasing w.r.t. $\varrho \in [0, R-1)$. The following numerical plots of v_0 support our conjecture.

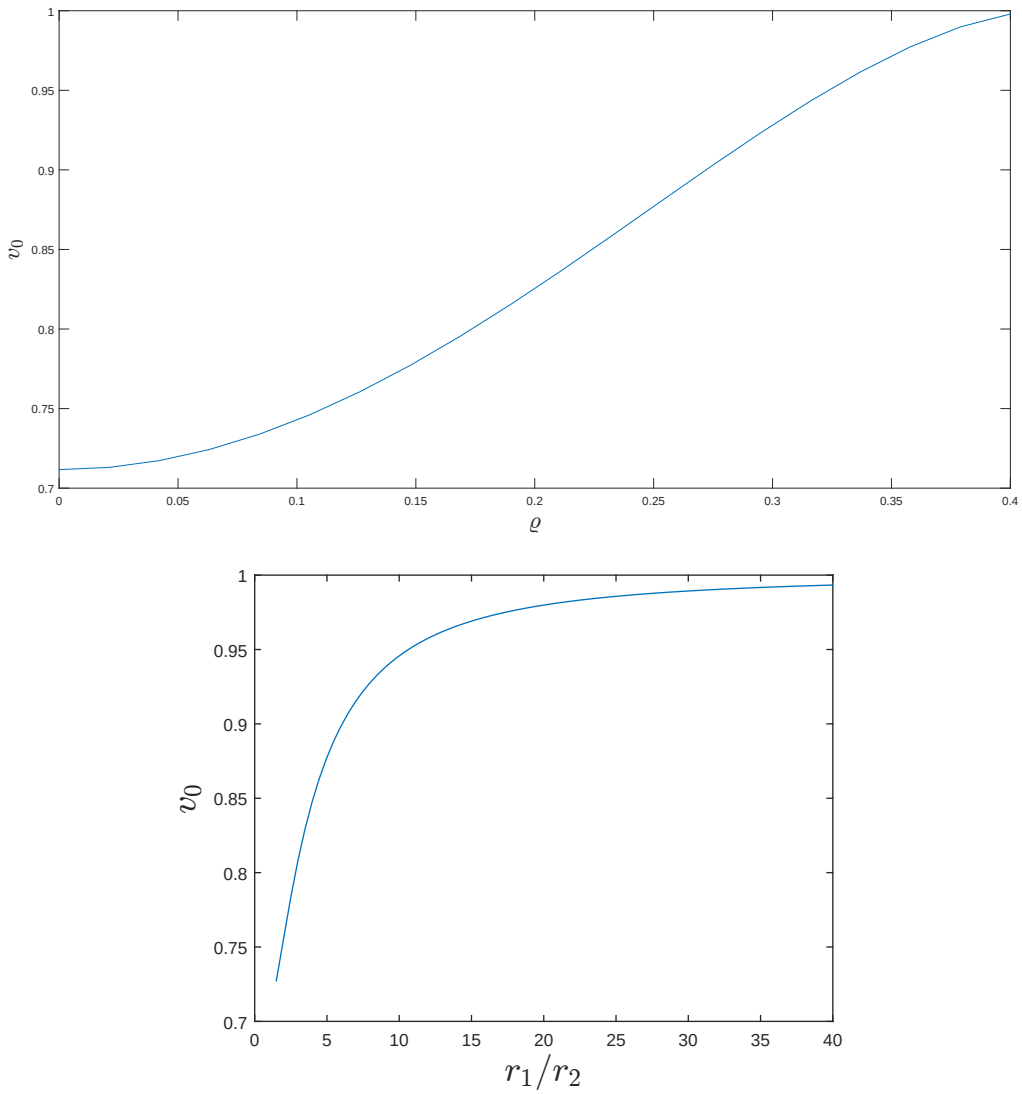


Figure 6.7: Reduced volume of cyclides $i_\varrho(T_{\sqrt{2}})$

By Theorem 6.0.4, the shape space of Clifford tori (6.2) is

$$\left\{ i_\varrho(T_{\sqrt{2}}) : \varrho \in [0, \sqrt{2}-1) \right\} \Big/ \mathrm{Hom}(3). \tag{6.32}$$

Each Clifford torus is uniquely determined by ϱ and it reduce volume increases w.r.t. ϱ. Therefore, the solution for the Canham problem is unique, when $v_0 \in [v_{SC}, 1)$.

List of References

[1] G. Arden. *Approximation properties of subdivision surfaces*. PhD thesis, University of Washington, Department of Mathematics, 2001. Available at http://grail.cs.washington.edu/projects/scanning/.

[2] David E. Blair. *Inversion theory and conformal mapping*. Student mathematical library. Providence, R.I. American Mathematical Society, 2000.

[3] W. Blaschke. *Vorlesungen ber Differentialgeometrie III*. Springer, 1929.

[4] A. I. Bobenko and P. Schröder. Discrete Willmore flow. In *In Eurographics Symposium on Geometry Processing*, pages 101–110, 2005.

[5] A. I. Bobenko and Y. B. Suris. *Discrete differential geometry*, volume 98 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2008.

[6] W. Boehm. On cyclides in geometric modeling. *Computer Aided Geometric Design*, 7(14):243 – 255, 1990.

[7] A. Bonito, R. H. Nochetto, and M. S. Pauletti. Parametric FEM for geometric biomembranes. *J. Comput. Phys.*, 229(9):3171–3188, 2010.

[8] K. A. Brakke. The surface evolver. *Experimental Mathematics*, 1(2):141–165, 1992.

[9] P. B. Canham. The minimum energy of bending as a possible explanation of the biconcave shape of the human red blood cell. *Journal of Theoretical Biology*, 26(1):61–76, 1970.

[10] V. Chandru, D. Dutta, and C. M. Hoffmann. On the geometry of dupin cyclides. *The Visual Computer*, 5(5):277–290, 1989.

[11] B. Y. Chen. An invariant of conformal mappings. *Proc. Amer. Math. Soc.*, 40:563–564, 1973.

[12] J. A. Cottrell, T. J. R. Hughes, and Y. Bazilevs. *Isogeometric Analysis: Toward integration of CAD and FEA*. Wiley, 2009.

[13] G. R. Cowper. Gaussian quadrature formulas for triangles. *International Journal for Numerical Methods in Engineering*, 7(3):405–408, 1973.

[14] C. de Boor, K. Höllig, and S. Riemenschneider. *Box Splines*. Springer-Verlag, 1993.

[15] E. A. Evans. Bending resistance and chemically induced moments in membrane bilayers. *Biophysical Journal*, 14(12):923 – 931, 1974.

[16] S. Grundel, J. Chen, and T. P.-Y. Yu. A flexible C^2 subdivision scheme for genus 0 surfaces. Preprints, August 2014.

[17] W. Helfich. Elastic properties of lipid bilayers: Theory and possible experiments. *Z. Naturforsch C*, 28(11):693–703, 1973.

[18] L. Hsu, R. Kusner, and J. Sullivan. Minimizing the squared mean curvature integral for surfaces in space forms. *Experimental Mathematics*, 1(3):191–207, 1992.

[19] T. J. R. Hughes, J. A. Cottrell, and Y. Bazilevs. Isogeometric analysis: CAD, finite elements, NURBS, exact geometry and mesh refinement. *Computer methods in applied mechanics and engineering*, 194(39-41):4135–4195, 2005.

[20] F. Jülicher, U. Seifert, and R. Lipowsky. Conformal degeneracy and conformal diffusion of vesicles. *Physical review letters*, 71(3):452–455, 1993.

[21] L. G. A. Keller, A. Mondino, and T. Rivière. Embedded surfaces of arbitrary genus minimizing the Willmore energy under isoperimetric constraint. *Archive for Rational Mechanics and Analysis*, pages 1–38, 2013.

[22] R. Lipowsky. The conformation of membranes. *Nature*, 349:475–481, 1991.

[23] C. T. Loop. Smooth subdivision surfaces based on triangles. Master's thesis, Department of Mathematics, University of Utah, 1987.

[24] F. C. Marques and A. Neves. Min-max theory and the Willmore conjecture. *Annals of Mathematics*, 179(2):683–782, 2013.

[25] J. C. Maxwell. On the cyclide. *Quarterly Journal of Pure and Applied Mathematics*, (9):111–126, 1868.

[26] X. Michalet and D. Bensimon. Observation of stable shapes and conformal diffusion in genus 2 vesicles. *Science*, 269(5224):666–8, 1995.

[27] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization, second edition*. World Scientific, 2006.

[28] J. Peters and U. Reif. *Subdivision Surfaces*. Springer-Verlag, Berlin, Heidelberg, 2008.

[29] H. Prautzsch and U. Reif. Degree estimates for C^k-piecewise polynomial subdivision surfaces. *Advances in Computational Mathematics*, 10(2):209–217, 1999.

[30] U. Reif. A degree estimate for subdivision surfaces of higher regularity. *Proceedings of the American Mathematical Society*, 124(7):153–174, 1996.

[31] U. Reif and P. Schröder. Curvature integrability of subdivision surfaces. *Advances in Computational Mathematics*, 14(2):157–174, 2001.

[32] J. Schygulla. Willmore minimizers with prescribed isoperimetric ratio. *Archive for Rational Mechanics and Analysis*, 203(3):901–941, 2012.

[33] U. Seifert. Configurations of fluid membranes and vesicles. *Advances in Physics*, 46(1):13–137, 1997.

[34] L. Simon. Existence of surfaces minimizing the Willmore functional. *Communications in Analysis and Geometry*, 1(2):281–326, 1993.

[35] J. Stam. Exact evaluation of Catmull-Clark subdivision surfaces at arbitrary parameter values. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '98, pages 395–404, New York, NY, USA, 1998. ACM.

[36] J. Stam. Evaluation of Loop subdivision surfaces. In *SIGGRAPH'99 Course Notes*, 1999.

[37] J. Warren and H. Weimer. *Subdivision Methods for Geometric Design: A Constructive Approach*. Morgan Kaufmann, 2001.

[38] D. Zorin and P. Schröder. Subdivision for Modeling and Animation. Technical report, SIGGRAPH 2000, 2000. Course Notes.